

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 1832.29 seconds
(without alignments)
- 11293.030 Million cell updates/sec

Title: US-09-105-117k-1_COPY_1016_1726

Perfect score: 711

Sequence: 1 atggtgatcatggaatctt.....aactgatgttgatgggttag 711

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_mus.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	711	100.0	822	6	AX063771	AX063771 Sequence
2	711	100.0	822	6	AX244059	AX244059 Sequence
3	711	100.0	2374	1	CGLXSEG	X96471 C.glutamicu
4	711	100.0	2374	6	AP93933	AP93933 Sequence 2
5	711	100.0	333150	1	AP005277	AP005277 Coryneb
6	711	100.0	349980	6	AX127147	AX127147 Sequence
7	708	99.6	708	6	AX123539	AX123539 Sequence
8	317.6	44.7	1771	1	AB083133	AB083133 Coryneb
9	84.2	11.8	190050	1	AL646059	AL646059 Ralstonia
10	65.6	9.2	33285	1	SC5F8	AL357613 Streptomy
11	63.4	8.9	10542	1	AE004852	AE004852 Pseudomon
12	62.8	8.8	24336	1	AE008841	AE008841 Salmonell
13	62	8.7	10701	1	AE005522	AE005522 Escherich
14	62	8.7	286658	1	AP002563	AP002563 Escherich
15	61.6	8.7	4033	1	ASU65741	U65741 Aeromonas s
16	61.2	8.6	230050	1	AL627277	AL627277 Salmonell
17	60.4	8.5	636	6	AX030085	AX030085 Sequence
18	60.4	8.5	8029	1	E49392	E49392 Process for
19	60.4	8.5	10362	1	ECFDPAGK	AE000375 Escherich
20	60.4	8.5	141744	1	ECU28377	U28377 Escherichia
21	60.4	8.4	213050	1	AL646067	AL646067 Ralstonia
22	59.6	8.4	10747	1	AE012410	AE012410 Xanthomon
23	57	8.0	14844	1	MEU34849	U34849 Mycobacteri
24	57	8.0	14869	1	AE007056	AE007056 Mycobacte
25	57	8.0	15239	6	I86263	I86263 Sequence 17
26	57	8.0	38500	1	MTCY39	MTCY39 Mycobacteri
27	56.4	7.9	11139	1	AE009057	AE009057 Agrobacte
28	56.4	7.9	12305	1	AE008023	AE008023 Agrobacte
29	56.2	7.9	14029	1	AE006952	AE006952 Mycobacte
30	56.2	7.9	37218	1	MTCY20G9	MTCY20G9 Mycobacte
31	56.2	7.9	10333	1	AE013931	AE013931 Yersinia
32	56	7.9	208050	1	AJ414145	AJ414145 Yersinia
33	55.4	7.8	10393	1	AE011957	AE011957 Xanthomon
34	55.4	7.4	12070	1	AB011413	AB011413 Streptomy
35	52.8	7.4	14135	1	AE009566	AE009566 Brucella
36	52.8	7.1	10698	1	AE004134	AE004134 Vibrio ch
37	50.2	6.9	2878	1	AHAYGENE	X89469 Aeromonas h
38	48.8	6.5	115666	2	AC105744	AC105744 Oryza sat
39	46	6.2	125020	9	AF429315	AF429315 Homo sapi
40	43.8	6.0	125020	9	AF429315	AF429315 Homo sapi
41	42.8	6.0	125020	9	AF429315	AF429315 Homo sapi
42	42.6	6.0	970	1	AB032557	AB032557 Streptomy
43	42.6	6.0	42210	1	SC1C2	AL031124 Streptomy
44	42.4	6.0	168398	2	AC097204	AC097204 Rattus no
45	42.2	5.9	149437	2	AP005610	AP005610 Oryza sat

ALIGNMENTS

RESULT 1
AX063771
LOCUS AX063771 822 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 53 from Patent WO0100843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 822)
REFERENCE
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway

JOURNAL	proteins		Patent: WO 0100843-A 53 04-JAN-2001;	
	BASF AKTIENGESELLSCHAFT (DE)		Location/Qualifiers	
	FEATURES	source	1. .822	
		CDS	/organism="Corynebacterium glutamicum"	
FEATURES	source	CDS	/db_xref="taxon:1718"	
			101. .802	
			/note="RXA01394"	
			/codon_start=1	
FEATURES	source	CDS	/transl_table=11	
			/protein_id="CAC25127.1"	
			/db_xref="GI:12541484"	
			/translation="MEIFITGLLGALLSIGPQNVLVIKQIKREGLIANVLVCLL	
FEATURES	source	CDS	SDVFLFIAGTGLVDLLSNAPIVLDIRWGGIAYLLWFAVMAKADKNTKVEAPQIIIE	
			ETEPVDDTPLGGSAVATDRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDA	
			FVFVGGAQYGDTRWIFAAAGFAASLIWFLPLVGFGAALSRPLSSPKVWRWVNV	
			AVVMTALAIKMLMG"	
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g
BASE COUNT	ORIGIN	167 a	192 c	246 g

```

QY 481 TTGACGCGTTTGTGTTATCGCGCGTGGCGGCAATACGGCAGACCGGACGGTGG 540
|
|
|
Db 572 TTGACGCGTTTGTGTTATCGCGCGTGGCGGCAATACGGCAGACCGGACGGTGG 631
|
|
|
QY 541 ATTTTCGCGCTGGCGGTTGGCGGCAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 600
|
|
|
Db 632 ATTTTCGCGCTGGCGGTTGGCGGCAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 691
|
|
|
QY 601 GCAGCAGCATTGTCACGCGCGTGTCCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 660
|
|
|
Db 692 GCAGCAGCATTGTCACGCGCGTGTCCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 751
|
|
|
QY 661 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 711
|
|
|
Db 752 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 802
|
|
|

RESULT 3
CGLYSEG
LOCUS
DEFINITION
ACCESSION X96471
VERSION X96471.1 GI:1729753
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine exporter protein; Lysine governor.
CORNEBACTERIUM GLUTAMICUM.
SOURCE
ORGANISM
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 (bases 1 to 2374)
Vrljic.M., Sahm.H. and Eggeling,L.
A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum
Mol. Microbiol. 22 (5), 815-826 (1996)
97126810
MEDLINE
PUBMED 8971704
REFERENCE
2 (bases 1 to 2374)
Vrljic.M.M.
Direct Submission
AUTHORS
Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie
JOURNAL
1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
LOCATION/Qualifiers
1. .2374
/organism="Corynebacterium glutamicum"
/strain="R127"
/db_xref="taxon:1718"
complement (82. .954)
/gene="lysG"
complement (82. .954)
/gene="lysG"
/note="lysine governor"
/codon_start=1
/transl_table=11
/product="Lysine export regulator protein"
/protein_id="CAA65323.1"
/db_xref="GI:1729754"
/db_xref="SWISS-PROT:P94632"
/translation="MNPQLDLSLIDEGFEGASIALSISPSAVSQKVALBHHVG
RLVSRTPAKATAGEVLVQARKMVLLOAETKQLSGLAEIPLTIAINADSLSTW
PPPVNEVASGGATLRLLEDEATLSLLRAGDVLGAVTRFANPVAGCEVLELGTWR
HLATATSLRDYAVVDGKLDWAAMPVLRFGPKVDLDRDLDRVDPGVRRRVSVPS
AEGFGEAIRRLGLHGLLPETQAPMLKAGEVILLDEIPIDTPMTWQRWRESRLARL
TDVVDVAREGLRP"
1025. .1726
/gene="lysE"
1025. .1726
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="CAA65324.2"

```

```

/db_xref="GI:13397387"
/db_xref="SWISS-PROT:P94633"
/translation="MEIFITGLLGLSLISIGPONVLVIKQIKREGLIALLVLCIL
SDVFLFIAGHLGDLDSNAAPILVIDIRWNGGIAYLLFWFAAKADNTKVEAPQILIE
ETEPVDDPTPLGSAVATDRNRVRVEVDKORVWVKPMLMAIIVLTWLNPNAYLDA
FVFTIGVGAQYGDGTGRWIFPAAGAFASLIWFLPLVFGAAALSRLSPSPKVRWIVVV
AVVMTALAIKMLMG"
complement (1723. .2352)
/gene="orf3"
complement (1723. .2352)
/gene="orf3"
/codon_start=1
/transl_table=11
/protein_id="CAA65325.1"
/db_xref="GI:1729756"
/db_xref="SPTRMBL:P94634"
/translation="MYALRDIVASGKALYVIGSYGPFLTAAAEFMAEBCPLLIHQ
PSYIINRWVEEPPDDGNLLQSAANNGLVIAFSPLAQGLLTKYLDGIPGSRASQ
GKSLSEGMLNVNNDIMVRKLDIAQERGQSLAQMALAWLRQEGYGAADVTSALIGA
SSVEQLDNLSDLSLNNLEFSDAELEAIDEISHDAGINIWAKATDSKTREN"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 711; DB 1; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.8e-168;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
Db 1016 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 1075
QY 61 ATCGACCGCAGATGTACTGTGATTAACAAGAAATTAAGCCGAGGACTCATTTGCG 120
Db 1076 ATCGACCGCAGATGTACTGTGATTAACAAGAAATTAAGCCGAGGACTCATTTGCG 1135
QY 121 GTTCTTCTCGTGTGTTAAATTTCTGACGTCTTTTGTTCATCGCGGCACCTTCGGGGTT 180
Db 1136 GTTCTTCTCGTGTGTTAAATTTCTGACGTCTTTTGTTCATCGCGGCACCTTCGGGGTT 1195
QY 181 GATCTTTTGCCAAATGCGGCGCGATCGTGTGATTAATATGCTGGGTGGGATCGCT 240
Db 1196 GATCTTTTGCCAAATGCGGCGCGATCGTGTGATTAATATGCTGGGTGGGATCGCT 1255
QY 241 TACCTGTTATGTTTGGTTCGCGTCATGCGAGCAAGAACCCCATGACAAACAGGTGGAAGCG 300
Db 1256 TACCTGTTATGTTTGGTTCGCGTCATGCGAGCAAGAACCCCATGACAAACAGGTGGAAGCG 1315
QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCTCCGATGACACGCTTTGGGCGGTTTCG 360
Db 1316 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCTCCGATGACACGCTTTGGGCGGTTTCG 1375
QY 361 CGGTTGGCCTGACACGCGCAACCGGTTGGGTGGGTGAGTGAGCTGATAGCAGCGG 420
Db 1376 CGGTTGGCCTGACACGCGCAACCGGTTGGGTGGGTGAGTGAGCTGATAGCAGCGG 1435
QY 421 GTTTGGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGTTGAACCCGAATGCCGTAT 480
Db 1436 GTTTGGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGTTGAACCCGAATGCCGTAT 1495
QY 481 TTGACGCGTTTGTGTTTATCGCGCGCTGGCGCGCAATACCGCGCAACCGGACGGTGG 540
Db 1496 TTGACGCGTTTGTGTTTATCGCGCGCTGGCGCGCAATACCGCGCAACCGGACGGTGG 1555
QY 541 ATTTTCGCGCTGGCGGTTTCGGCGCAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 600
Db 1556 ATTTTCGCGCTGGCGGTTTCGGCGCAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 1615
QY 601 GCAGCAGCATTGTCACGCGCGTGTCCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 660
Db 1616 GCAGCAGCATTGTCACGCGCGTGTCCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 1675
QY 661 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 711
|
|
|

```

Db 1676 GTGGCAGTTGTGATGACCGCATTGGCCATCAAACTGATGTTTGTATGGGTTAG 1726

RESULT 4
LOCUS A93933 2374 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 2 from Patent WO9723597.
ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 2374)
REFERENCE Vrljic.M. and Eggeing.L.
AUTHORS PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
TITLE ACTIVITY OF EXPORT CARRIERS
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)
FEATURES
source Location/Qualifiers
1. 2374
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 711; DB 6; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.8e-168;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 60
Db 1016 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 1075
QY 61 ATCGGACCCGACAACTGCTGTGATTAACAAGAAATTAACGCGAAGGACTCAATTCG 120
Db 1076 ATCGGACCCGACAACTGCTGTGATTAACAAGAAATTAACGCGAAGGACTCAATTCG 1135
QY 121 GTTCTTCTGCTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGCGGTT 180
Db 1136 GTTCTTCTGCTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGCGGTT 1195
QY 181 GATCTTTTGTCAATGCCGCGCATGCTGCTCGATATATCGCTGGGGTGGATCGCT 240
Db 1196 GATCTTTTGTCAATGCCGCGCATGCTGCTCGATATATCGCTGGGGTGGATCGCT 1255
QY 241 TACCTGTTATGTTTGGCTGATGCGGAAAGAGCGCATGACAAACAAGTGAAGCG 300
Db 1256 TACCTGTTATGTTTGGCTGATGCGGAAAGAGCGCATGACAAACAAGTGAAGCG 1315
QY 301 CCACAGATCATTAAGAAACAGAACCAACCGTGCCCGATGACACGCTTTGGCGGCTTCG 360
Db 1316 CCACAGATCATTAAGAAACAGAACCAACCGTGCCCGATGACACGCTTTGGCGGCTTCG 1375
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTCGCGGTGGAGGTGAGCTGATGAAGCAGCG 420
Db 1376 GCGGTGGCCACTGACACGCGCAACCGGTCGCGGTGGAGGTGAGCTGATGAAGCAGCG 1435
QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGCAACCCGAATCGGTAT 480
Db 1436 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGCAACCCGAATCGGTAT 1495
QY 481 TTGAGCGGTTTGTGTTTATCGCGCGCTGCGCGCGCAATACCGCGACACCGGACGGTGG 540
Db 1496 TTGAGCGGTTTGTGTTTATCGCGCGCTGCGCGCGCAATACCGCGACACCGGACGGTGG 1555
QY 541 ATTTTCGCGCTGCGCGCTTTCGCGCAAGCTGATCTGTTCCCGCTGGTGGGTTTCGCG 600
Db 1556 ATTTTCGCGCTGCGCGCTTTCGCGCAAGCTGATCTGTTCCCGCTGGTGGGTTTCGCG 1615
QY 601 GCAGCAGCATTTGTTCACGCGCGCTGTTCAGCCGCCAGGCTGTGGCTGGATCAACGCTGTC 660

Db 1616 GCACACATTGTACCCCGCTGTCCAGCCCAAGGTGTGGCGTGCATCAACGTCGTC 1675

QY 661 GTGGCAGTTGTGATGACCGCATTGGCCATCAAACTGATGTTTGTATGGGTTAG 711

Db 1676 GTGGCAGTTGTGATGACCGCATTGGCCATCAAACTGATGTTTGTATGGGTTAG 1726

RESULT 5
LOCUS AP005277/c 333150 bp DNA linear BCT 08-AUG-2002
DEFINITION Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 4/10
ACCESSION AP005277 BA000036
VERSION AP005277.1 GI:21323710
KEYWORDS
SOURCE Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.
ORGANISM Corynebacterium glutamicum ATCC 13032
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
1
REFERENCE Nakagawa,S.
AUTHORS Complete genomic sequence of Corynebacterium glutamicum ATCC 13032
TITLE Unpublished
JOURNAL 2 (bases 1 to 333150)
REFERENCE Nakagawa,S.
AUTHORS Direct Submission
TITLE Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.
JOURNAL Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida,
Tokyo 194-8533, Japan (E-mail:snakagawa@xanagen.com,
Tel:81-44-829-3031, Fax:81-44-813-1651)
COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo
Co. Ltd. And Kitasato University.
FEATURES
source Location/Qualifiers
1. 333150
/organism="Corynebacterium glutamicum ATCC 13032"
/strain="ATCC 13032"
/db_xref="taxon:196627"
/notes="ATCC 13032"
complement (63..1292)
/gene="Cgl10944"
complement (63..1292)
/gene="Cgl10944"
/codon_start=1
/transl_table=11
/product="Hypothetical membrane protein"
/protein_id="BAB98337.1"
/db_xref="GI:21323711"
/translation="MTNPTERNARBLIHWGLQNTQDQIVAAKTIVLPWLLQAAGP
FLALVPIREAGMLPQRAITGWLRSRQVWVIGNSQFVSAIGVAAALFLRG
WALGITVIALLAALSIFRSMCSIASKDVQKVIKRGKGLVTGRATVIGVGMVLVAGL
AIAFLGSHSPTRVLAAVITASSFWLFAIYKARIEYKAPATPKNAPSANPWRRCI
AALDDKAFRRFVLVRSMLVTALSTAFIYKARIEYKAPATPKNAPSANPWRRCI
RISGITSWSKNNMAGGALFGSIVILVLSAFAPAOINTLVFPLSLITLAHTA
IRVARTYVMDMAEGDQRTFYVADANTLMGVLLIIVGALSGLTAIFGNEAALLFLAAI
GLLGITSARGLKEVSAG"
1382..2917
/gene="Cgl10945"
1382..2917
/gene="Cgl10945"
/codon_start=1
/transl_table=11
/product="Putative multicopper oxidases"
/protein_id="BAB98338.1"
/db_xref="GI:21323712"
/translation="MPENLRRTFFKAGVLAATVGAQVLVACSSDDVRVGGSPRTL
PIPADLGTREGSSVHFALEAQTGESQILPDVTTKTGNGTHLPTLVVKGDDVHV
DVINLDMETVHWHGKLPALADGPHSPFGQTWSPTVTVANDAAATLWYRPHTHG
LTGLHAYRGLAGMLIIVEDATDKLDLPREYGVDDIPLVLMHMRFLPDGSLDEBDLP
GLGDTTANGITNAHFATIRVRFRVLNGSNMRYNLAFSDFTRTFQVTFASGSLD
EPQDRTTLAIPGERWEIIVVELEPDEDVLESVGFEDYGVDPDEFVDFPGMDSFOL
LTTTPSDDAAQAAPALPGVLVKSTEPDVIDATERTFMNTFSINDLQMDMQRVDDVID
HDQPEWIVTNDNSDPHNFHVDAREKVLKFESTDVDELFDNDGKDKTGLVLPQATATL


```

AVEFGHYDPQWPMYHCHMLYHEDQGMGQFVIVPGDEPAAVLGSGTSSIDSAGS
RAH"
gene
3063. .3983
/ gene="Cg10946"
CDS
3063. .3983
/ gene="Cg10946"
/ note="PF00005:ABC transporter
TIGR00960:3a0501802: Type II (General) Secretory Pathway
(IISP) Family protein
TIGR00968:3a0106801: sulfate transport system permease
protein
TIGR01166:cbiO: cobalt transport protein ATP-binding
subunit
TIGR01184:ntrCD: nitrate transport ATP-binding subunits C
and D
TIGR01186:prov: glycine betaine/L-proline transport ATP
binding subunit
TIGR01187:potA: spermidine/putrescine ABC transporter
ATP-binding subunit
TIGR01188:draA: daunorubicin resistance ABC transporter
ATP-binding subunit
TIGR01189:ccmA: heme exporter protein CcmA"
/ codon_start=1
/ transl_table=11
/ product="ABC-type transporter, ATPase component"
/ protein_id="BAB98339.1"
/ db_xref="GI:21323713"
/ translation="MTTPAIDVTLVRYGDTAVKGLNFHVORGEVGLLTNGAG
KTSLEVEGLSASGGVIRISGLDPVADRAILPELIGIMLQSGLSQTLVAETMD
WHGCTYPRALKVDLVDLLHRENVKVALSGEGEQRDLACALLGDPSTFLFDEPT
TGLDPESRHWQLLDLQKQVMTMLTHYLEAEFLCDRIAIMNAGEIAVEGTLDL
LVAREKSIISFVLGGQVGVLSGAEIIRDNHVRVRIATTLQOHTLEILTAAETGI
ALEGFAAKPATLESVFMDIASLENTSLQTA"
4003. .4836
/ gene="Cg10947"
CDS
4003. .4836
/ gene="Cg10947"
/ codon_start=1
/ transl_table=11
/ product="ABC-type transporter, permease components"
/ protein_id="BAB98340.1"
/ db_xref="GI:21323714"
/ translation="MTTSHTARGLOHATPERRKTSFFKTSLFKAEMLQPRNKTLFPM
ATVPFVGLPLFLIINGGAESAESDFVWYTLFFVQFVTLSMATRRDERVLKR
LRTGEARDIDIGAIICPGALLTLFTVWIPLMLVGAPAPINLPIVAVLIGLL
CSALATSGFTNAEAQNTSMPEVFLMAGLGSIRPFVGDSSIVADILAYTPPAIS
DLVQIWAGATFADSVGVEAANFAGIFQDMLEIFLIGILAATAAANAANRYMRWDSY
R"
4879. .6138
/ gene="Cg10948"
CDS
4879. .6138
/ gene="Cg10948"
/ EC_number="2.7.3.-"
/ codon_start=1
/ transl_table=11
/ product="Two-component system, sensory transduction
histidine kinases"
/ protein_id="BAB98341.1"
/ db_xref="GI:21323715"
/ translation="MNKDFWTAGTWARFSGVSLASPVTPAPLNSWRRLPNLAKYTL
YTRVSLQAI PVLLSAYFLGIVANAGTINPFSFVMLGFSVILLITVTVVEYQPSLMS
HPRSVPQFFFTGLVNLGVVSVLQIPGLNMSDNRATALI FTLCVPLLSIAYI
PMWNYRVLIAWSAVLWNTSTTDDYLSALWVLPMLAGTVLSVMTVDVMEKVERS
RELASLRVTEERLPAQELHDTLQGLHAMS VKSELALAKKGGDRRLNEURELQK
LRTSMEMRDVSGYRTVNLATIEGAKSLLDADAHILHSVIGTTSQVSPAHRELCAW
LVREATTNLRHSDDATATLSTSTVYRMDNNGVYKDIIGRLSGLSALRSRAGMTL
IVSRDDQFSVRMLINAPANTPAEKA"
6139. .6750
/ gene="Cg10949"
CDS
6139. .6750
/ gene="Cg10949"
/ note="PF00072:Response regulator receiver domain
PF00196:Bacterial regulatory proteins, luxR family"

```

```

/ codon_start=1
/ transl_table=11
/ product="Two-component system, response regulators
consisting of a CheY-like receiver domain and a HTH
DNA-binding domain"
/ protein_id="BAB98342.1"
/ db_xref="GI:21323716"
/ translation="MISISTADDEALIASLATLLSLEPDLVDPRTAGSGEELIETWA
DPNRRYDVCVLQDLGSDIGDITATRLMETTPDLAVLIVTSHARPQLKRALAGVLG
FLKTSIADBPATAIRVHAGRYIDPELAAMTISAGESPLTNREVEVLELACQGLSA
EEIATAHAPAGTTRNYLQAMTKVGAQNRFEAFTRARELGWL"
complement (6781. .6987)
/ gene="Cg10950"
CDS
complement (6781. .6987)
/ gene="Cg10950"
/ codon_start=1
/ transl_table=11
/ product="Hypothetical protein"
/ protein_id="BAB98343.1"
/ db_xref="GI:21323717"
/ translation="MKHYMSWYFAGPLNTPHTKVFYWFMDIVSVVNHWFEMSTNNVN
MTGLPRETFVPAFVGSTVLAFLSF"
7048. .8205
/ gene="Cg10951"
CDS
7048. .8205
/ gene="Cg10951"
/ codon_start=1
/ transl_table=11
/ product="tRNA-processing ribonuclease BN"
/ protein_id="BAB98344.1"
/ db_xref="GI:21323718"
/ translation="MTAWLCSPSPGVQDEVMMGVVQPEHLDATLIAADPHGNEN
SGDKERLNFQCKYALNCTVRDVPDGLDLDAALLTFSSILSIAPVALLGYVILIF
LASDTEILNLVRDEVNVPDQSHVNGVIDSIAGSAAGQGVAVGVITALTSS
AYVRAFSCANAVYGRSEGLTIKRWALLFLNALLGLIIILVSWNLNETLVMGIF
APTAEPHLTNLVSFLTDRFMTIWIWRFVIVGLVIMFVATLWYAPNAPKFWRL
SLGSFLAIVGILLAGVLFNYFTLFAFVSYGAVGSLLAFFVIALWVFNICLIIGLKID
VEISRAKOLQAGMPAEDYSLVPRPSIEKAKMKQKQRLMDQAAARESN"
8672. .10114
/ gene="Cg10952"
CDS
8672. .10114
/ gene="Cg10952"
/ notes="PF00005:ABC transporter
PF00664:ABC transporter transmembrane region."
/ codon_start=1
/ transl_table=11
/ product="ABC-type multidrug/protein/lipid transport
system, ATPase component"
/ protein_id="BAB98345.1"
/ db_xref="GI:21323719"
/ translation="MWRTHTASSILRRMRQRKVARFAGFLGMWQLSEALVPIAI
GLIVDHAVLTKDLRLVGLVAVFLVFLVSFYSFSGSALNRANRVFESHARVEVAD
HAKNLDPRLNPFGEVMSRTADSDSTIFQIGTGVSAATGFLGAATYLLISDLV
GLLVLVLPITISGVVALASKGISKRSTQOEKLAESQAOSDMMGLRVTKAIGSRW
AVTKPEKSAQASARAADVTAVASGVAGVIGELSLAVNLAALLACWRTVTCGLGPQ
LAIACVAVYLSPEIRLLNSINASAIAGAAERVANFNLDESQAQYSSSETINDGE
FLIVPPASTLPHGDNLATPHADIFEGTLKRSNTSMNHEDNVPIDQVIRASGLTDI
IEVDGLDAPVDTGTSNLGQQRVALRALHADAEVLVMDPTSAVDSVTEVSIQAQ
IKQLRAGKTTIVVSSSPAFYNLADRIVSHV"
10107. .11843
/ gene="Cg10953"
CDS
10107. .11843
/ gene="Cg10953"
/ notes="PF00005:ABC transporter
Query Match 100.0%; Score 711; DB 1; Length 333150;
Best Local Similarity 100.0%; Pred. No. 5.5e-168;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTGTATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
|||||
Db 329003 ATGTGTATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 328944
|||||

```

QY 61 ATCGACCGCAGAAATGATGCTGATTAACCAAGAAATTAAGCGAAGGACTCATTCGG 120
 |||||
 Db 328943 ATCGACCGCAGAAATGATGCTGATTAACCAAGAAATTAAGCGAAGGACTCATTCGG 328884
 QY 121 GTTCTTCTCGTGTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGGCACCTTTGGGCGTT 180
 |||||
 Db 328883 GTTCTTCTCGTGTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGGCACCTTTGGGCGTT 328824
 QY 181 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATATATGCTGTGGGTGGCATCGCT 240
 |||||
 Db 328823 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATATATGCTGTGGGTGGCATCGCT 328764
 QY 241 TACCTGTTATGTTTGGCGTATGCGAGCAAGCAAGCGCATGACAAACAAGTGAAGCG 300
 |||||
 Db 328763 TACCTGTTATGTTTGGCGTATGCGAGCAAGCAAGCGCATGACAAACAAGTGAAGCG 328704
 QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTTCG 360
 |||||
 Db 328703 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTTCG 328644
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGAGGTGAGCTGATGAAGCAGCGG 420
 |||||
 Db 328643 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGAGGTGAGCTGATGAAGCAGCGG 328584
 QY 421 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGTGCTGACCTGTTGAACCCGAATCGGTAT 480
 |||||
 Db 328583 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGTGCTGACCTGTTGAACCCGAATCGGTAT 328524
 QY 481 TTGGACGCGTTTGTGTTTATCGCGCGCGTTCGGGCGCAATACGGCGACACCGGACGTTGG 540
 |||||
 Db 328523 TTGGACGCGTTTGTGTTTATCGCGCGCGTTCGGGCGCAATACGGCGACACCGGACGTTGG 328464
 QY 541 ATTTTCGCGTGGCGGCTTCGGCGCAAGCTGATCTGTTCCCGTGGTGGTTTCGGC 600
 |||||
 Db 328463 ATTTTCGCGTGGCGGCTTCGGCGCAAGCTGATCTGTTCCCGTGGTGGTTTCGGC 328404
 QY 601 GCAGCAGCATTTGACGCGCGCTTCAGCCCGCAAGGTTGCGCTGATCAACGTCGTC 660
 |||||
 Db 328403 GCAGCAGCATTTGACGCGCGCTTCAGCCCGCAAGGTTGCGCTGATCAACGTCGTC 328344
 QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 711
 |||||
 Db 328343 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 328293

RESULT 6
 AX127147/c AX127147 349980 bp DNA linear PAT 11-MAY-2001
 LOCUS
 DEFINITION Sequence 7063 from Patent EP1108790.
 ACCESSION AX127147 AX114121
 VERSION AX127147.1 GI:14041135
 KEYWORDS
 SOURCE
 ORGANISM
 Corynebacterium glutamicum.
 Corynebacterium glutamicum
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 Corynebacterium.

REFERENCE 1 (bases 1 to 349980)
 AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
 TITLE Novel polynucleotides
 JOURNAL Patent: EP 1108790-A 7063 20-JUN-2001;
 KYOWA HAKKO KOGYO CO., LTD. (JP)
 FEATURES
 Location/Qualifiers
 source
 1. 349980
 /organism="Corynebacterium glutamicum"
 /db_xref="taxon:1718"
 /note="Seq 1 to long (3.309.400) split in 11, seq 7063
 1.200.001 1.549.980"
 BASE COUNT 80727 a 91049 c 97618 g 80586 t
 ORIGIN

Query Match 100.0%; Score 711; DB 6; Length 349980;

Best Local Similarity 100.0%; Pred. No. 5.5e-168;
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTGTATCATGGAATCTTTCTATACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 60
 |||||
 Db 128953 ATGTGTATCATGGAATCTTTCTATACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 128894
 QY 61 ATCGGACCGCAGAAATGATGCTGATTAACCAAGAAATTAAGCGCGAAGACTCATTCGG 120
 |||||
 Db 128893 ATCGGACCGCAGAAATGATGCTGATTAACCAAGAAATTAAGCGCGAAGACTCATTCGG 128834
 QY 121 GTTCTTCTCGTGTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGGCACCTTTGGGCGTT 180
 |||||
 Db 128833 GTTCTTCTCGTGTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGGCACCTTTGGGCGTT 128774
 QY 181 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATATATGCTGTGGGTGGCATCGCT 240
 |||||
 Db 128773 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATATATGCTGTGGGTGGCATCGCT 128714
 QY 241 TACCTGTTATGTTTGGCGTATGCGAGCAAGCAAGCGCATGACAAACAAGTGAAGCG 300
 |||||
 Db 128713 TACCTGTTATGTTTGGCGTATGCGAGCAAGCAAGCGCATGACAAACAAGTGAAGCG 128654
 QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTTCG 360
 |||||
 Db 128653 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTTCG 128594
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGAGGTGAGCTGATGAAGCAGCGG 420
 |||||
 Db 128593 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGAGGTGAGCTGATGAAGCAGCGG 128534
 QY 421 GTTTGGGTAAAGCCCATGTTGATGCGAATCGTGTGCTGACCTGTTGAACCCGAATCGGTAT 480
 |||||
 Db 128533 GTTTGGGTAAAGCCCATGTTGATGCGAATCGTGTGCTGACCTGTTGAACCCGAATCGGTAT 128474
 QY 481 TTGGACGCGTTTGTGTTTATCGCGCGCGTTCAGCCCGCAATACGGCGACACCGGACGTTGG 540
 |||||
 Db 128473 TTGGACGCGTTTGTGTTTATCGCGCGCGTTCAGCCCGCAATACGGCGACACCGGACGTTGG 128414
 QY 541 ATTTTCGCGTGGCGGCTTCGGCGCAAGCTGATCTGTTCCCGTGGTGGGTTTCGGC 600
 |||||
 Db 128413 ATTTTCGCGTGGCGGCTTCGGCGCAAGCTGATCTGTTCCCGTGGTGGGTTTCGGC 128354
 QY 601 GCAGCAGCATTTGACGCGCGCTTCAGCCCGCAAGGTTGCGCTGATCAACGTCGTC 660
 |||||
 Db 128353 GCAGCAGCATTTGACGCGCGCTTCAGCCCGCAAGGTTGCGCTGATCAACGTCGTC 128294
 QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 711
 |||||
 Db 128293 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 128243

RESULT 7
 AX123539 AX123539 708 bp DNA linear PAT 11-MAY-2001
 LOCUS
 DEFINITION Sequence 3455 from Patent EP1108790.
 ACCESSION AX123539
 VERSION AX123539.1 GI:14041027
 KEYWORDS
 SOURCE
 ORGANISM
 Corynebacterium glutamicum.
 Corynebacterium glutamicum
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 Corynebacterium.
 1 (bases 1 to 708)
 AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
 TITLE Novel polynucleotides
 JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;
 KYOWA HAKKO KOGYO CO., LTD. (JP)
 FEATURES
 Location/Qualifiers
 source
 1. 708
 /organism="Corynebacterium glutamicum"

```

BASE COUNT      134 a 173 c 221 g 180 t
ORIGIN
Query Match      99.6%; Score 708; DB 6; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.4e-167;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
DB 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
QY 61 ATCGGACCGGAGATGTAATGATGATTAACAAGGAATTAAGCGGAGGACTCAATTGCG 120
DB 61 ATCGGACCGGAGATGTAATGATGATTAACAAGGAATTAAGCGGAGGACTCAATTGCG 120
QY 121 GTTCTTCTCGTGTTTAATTTCTGAGCTCTTTTGTTCATCGCGGACCTTGGCGTT 180
DB 121 GTTCTTCTCGTGTTTAATTTCTGAGCTCTTTTGTTCATCGCGGACCTTGGCGTT 180
QY 181 GATCTTTTGTCCAAATCGCGCGGATCGTGTCTGATATTATGCTGGGTGGCATCGCT 240
DB 181 GATCTTTTGTCCAAATCGCGCGGATCGTGTCTGATATTATGCTGGGTGGCATCGCT 240
QY 241 TACCTGTTATGTTTCCGTCATGCGGAGAAAGCGCATGACAAACAAGGTGGAAGCG 300
DB 241 TACCTGTTATGTTTCCGTCATGCGGAGAAAGCGCATGACAAACAAGGTGGAAGCG 300
QY 301 CCACAGATCATTCAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTCC 360
DB 301 CCACAGATCATTCAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTCC 360
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGTGGAGTGCATTAAGCAGCG 420
DB 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGTGGAGTGCATTAAGCAGCG 420
QY 421 GTTGGGTTAAACCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGAATCGTAT 480
DB 421 GTTGGGTTAAACCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGAATCGTAT 480
QY 481 TTGGAGCGGTTGTTTATCGCGCGGTGCGGCGCAATACGGCGACACCGGAGGTTGG 540
DB 481 TTGGAGCGGTTGTTTATCGCGCGGTGCGGCGCAATACGGCGACACCGGAGGTTGG 540
QY 541 ATTTTCGCGCTGGCGGTTCGCGGCAAGCTGATCTGTTCCCGTGGTGGTTTCGCG 600
DB 541 ATTTTCGCGCTGGCGGTTCGCGGCAAGCTGATCTGTTCCCGTGGTGGTTTCGCG 600
QY 601 GCAGCAGATTTGTCACGCGCGTGTCCAGCCCAAGGTGTGCGGTGGATCAACGTGCTC 660
DB 601 GCAGCAGATTTGTCACGCGCGTGTCCAGCCCAAGGTGTGCGGTGGATCAACGTGCTC 660
QY 661 GTGGCAGTTGTCATGACCGCATTTGGCCATCAAACTGATGTTGATGGT 708
DB 661 GTGGCAGTTGTCATGACCGCATTTGGCCATCAAACTGATGTTGATGGT 708

RESULT 8
LOCUS AB083133
DEFINITION Corynebacterium efficiens lysG, lysE genes for Lysine export
transcriptional regulatory protein, Lysine exporter protein,
complete cds.
ACCESSION AB083133
VERSION AB083133.1 GI:20065738
KEYWORDS
SOURCE
ORGANISM Corynebacterium efficiens DNA.
Corynebacterium efficiens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1
AUTHORS Itaya,H., Kimura,E., Kawahara,Y. and Sugimoto,S.
TITLE lysG, lysE of Corynebacterium efficiens
```



```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD13913.1"
/db_xref="GI:17427395"
/translation="MTGPGSPASPLIGALSFGFGLASLIIVAIGQNAVYLRQGLRRE
YVLGVLLCALCDMALIALGVAGMGTLLI SAHPAMLTAVRWAGAAPFLIAYGARAFRAAW
RGAERLQARNQDKASHAQVLAASALALSLLNPHVYLDITVLLGATGGRVAMPANVAFAG
GAMCASILWFLSLGFGARLLRPVPFARPVAVRWVLDALIGAVMVAIALTLILMG"
complement(2924..3910)
/gene="ftsY"
/feature="RSC0386; RSC0354"
complement(2924..3910)
/gene="ftsY"
/function="cell processes; cell division"
/note="Product confidence : putative
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE CELL DIVISION PROTEIN"
/protein_id="CAD13914.1"
/db_xref="GI:17427396"
/translation="WPDTPAAPALDMQADDIEVTVPVVEQARKGMSRLRSLGSK
TSKSLTTLFVGKVDLEALFELETLALMADAGVDATEYLLDELRRVKAQRIETAEVG
KTAURLDLIELHLEKTMVLEGRDPTVIMTAGVNGAKTTSIGKLCCKHFYTGQSVL
LAAGDTFAARAEQVINGQRNVTVVAQESGDPAAVIFDVAARARGIDIVNADTA
GRLPQLHMEELKVRVIGKAMATAPHELTLLVIDANTGONALQVAKAFDGLG
LIVTKLDGTAKGGILAAIARQRPVPVYFVGVEQVEDLQPSAREFADALLG"
4124..5629
/gene="RSC0387"
/note="RSC0387"
4124..5629
/gene="RSC0387"
/EC_number="3.4.99.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE ZINC PROTEASE PROTEIN"
/protein_id="CAD13915.1"
/db_xref="GI:17427397"
/translation="WRSFPOSNSARTGLRWAGITLAVALLAGEAQAQAPAPMARQ
AAALTGAKVTSQADTVEYRIANGRLIVREDHAPTVAHMVVTHAGSIDEHNGTTGV
AHMLHEHMFKTRAVGEGFSRRVAALGGRENAMTDTFTMYFOQIEKSLADVMAL
ADRWANLQLTDEKFPENNVKERRMRIDDSRATVVEQMLAVLFNAAPVNPFTIG
PSDLTMTQDAQVHKVAPNNATVTVVDVNPDEVFQAOQYCKLOPHALPRY
AQDEPKGVGKRIWKAAPENPVYLVAKAPPLDEKVDVDPVALEVLAVLQYDNA
RLPNLLVKGKDEKLVADNAGYDGNRGPSPFLIDGVPADGHTTAEIQAOLRAQI
DRIARDGTEALRKRAQVVAQIYKRDVSFGQGMIEGMAEMTGLSRLDRLILEKI
KSVTPAQVQVAKTYFTEDNLVATLLPQPIDPNPKAPKVPVPGNREEGGLR"
5629..6972
/gene="RSC0388"
/note="RSC0388"
5629..6972
/gene="RSC0388"
/EC_number="3.4.99.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by FrameD"
/codon_start=1

```

```

/evidence=not experimental
/transl_table=11
/product="PUTATIVE ZINC PROTEASE-LIKE SIGNAL PEPTIDE
PROTEIN"
/protein_id="CAD13916.1"
/db_xref="GI:17427398"
/translation="MRAMSLTLTKTFLAGMLAAAHGALAALPIEHWTAAATGARVFFVP
SPGIPMLDINLVDVAGTRYPAAKVGSLASTAGMLDKGVAAGSTPARDAAATADAF
DVGASPSGGAGGDRTSLSRLTSLDPAERKPAVDLMAQIVNAAPVPAVLTRDKQRTVA

```

```

Query Match      11.8%; Score 84.2; DB 1; Length 190050;
Best Local Similarity 47.9%; Pred. No. 3.4e-10;
Matches 242; Conservative 0; Mismatches 263; Indels 0; Gaps 0;

Q2 207 CGTCTCGATATTATCGCTGGGTGGCATCGCTTACTTGTATGTTTCCGTCATGGC 266
      |||||
Db 2423 CGTCTCGCGCTGTGTCTGATCTCGCGCTGTGCGACATGGCGCTGATCGCGCTGGGCGT 2482
      |||||

Q2 267 AGCGAAAGACCCCATGACAAACAGGTGGAAGCGCCACAGATCATTTGAAGAAACAGAAC 326
      |||||
Db 2483 GGCGGGCATGGCACGCTGATCTCGCGCACCCGGCATGCTGACGGCGGTGCGCTGGGC 2542
      |||||

Q2 327 AACCTGCCCCGATGACACGCTTTTGGCGGTTCGGCGGTGGCCACTGACACGCGCAACCG 386
      |||||
Db 2543 CGGCGCGGCTTCTGCTGGCTTACGGCGCACGGGCGTTCGGGCGGCGTGGCGCGCGC 2602
      |||||

Q2 387 GGTGGGTGGAGTGAGCTCGATAGCAGCGGTTTGGTAAAGCCCATGTTTATGTCGCG 446
      |||||
Db 2603 CGAGCGGCTGCGAGCGCGCAACCGCGCAAGCGCTCGCATGCGCAGGTGCTGGCGTCCGC 2662
      |||||

Q2 447 AATCGTCTGACCTGTTGAACCCGAATGCTATTTCGACGCGTTCGTGTTTATTCGCGCG 506
      |||||
Db 2663 GCTGGCGTGTGCTGCTCAATCGCAGCTATCTCGACACCGTGTGCTGCTGGCGCGC 2722
      |||||

Q2 507 GTCGCGCGCAATACCGCGCACACCGCGGTGATTTTCGCGCTGCGCGCTTCGCGCGC 566
      |||||
Db 2723 CATCGCGCGCGCTACGCGATCGCGCAACGTGGCGTTCGCGCGCGCGCATGTGCGC 2782
      |||||

Q2 567 AAGCTGATCTGTTCCGCTGGTGGGTTTCGGCGCAGCAGCATTTGTCAGCCCGCTGTC 626
      |||||
Db 2783 GTCGATCTGTTGTTCTGCTGTGGGTTTCGGCGCAGCGCTGCTGGAGCGGTGTTTCGC 2842
      |||||

Q2 627 CAGCCCCAAGGTGCGCGCTGGATCAACGTCGTCGTCGCGCAGTTGTGATGACCGCATGGC 686
      |||||
Db 2843 CAGCGCGTGCCTGGCGGCTGCTCGATGCGCTGATCGCGCGGCTGATGTGGGCGATCGC 2902
      |||||

Q2 687 CATCAACTGATGTTGATGGTTAG 711
      |||||
Db 2903 GCTGACGCTGCTGATGGCGGTTAG 2927
      |||||

```

RESULT 10

```

SCSF8
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

```

SCSF8 33285 bp DNA linear BCT 12-MAY-2002

Streptomyces coelicolor cosmid 5F8.

AL357613 AL645882

AL357613.2 GI:20520783

amino acid decarboxylase; anti-sigma factor; integral membrane protein; LuxR-family regulator; LysR-family transcriptional regulator; membrane transport protein; oxidoreductase; regulatory protein; RNA polymerase sigma factor; serine/threonine protein kinase; stress-inducible protein; TetR-family transcriptional regulator; thioredoxin reductase; threonine dehydratase; two-component sensor histidine kinase.

Streptomyces coelicolor A3(2).

Streptomyces coelicolor A3(2).

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 33285)

Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J., Kinashi, H., and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

```

JOURNAL      Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE      97000351
PUBMED       8843436
REFERENCE    2 (bases 1 to 33285)
AUTHORS      Seeger,K.J. and Harris,D.
JOURNAL      Unpublished
REFERENCE    3 (bases 1 to 33285)
AUTHORS      Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
JOURNAL      Direct Submission
TITLE        Submitted (05-JUN-2000) Streptomyces coelicolor sequencing project,
              Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
              CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
              David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
              Colney, Norwich, Norfolk NR4 7UH, UK
              On May 9, 2002 this sequence version replaced gi:8347023.
COMMENT      Notes:
              Streptomyces coelicolor sequencing at The Sanger Centre is funded
              by the BBSRC and Beowulf Genomics
              Details of S. coelicolor sequencing at the Sanger Centre are
              available on the World Wide Web.
              (URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
              numbered using the following system eg SC7B7.01c. SC (S.
              coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
              strand).
              The more significant matches with motifs in the PROSITE database
              are also included but some of these may be fortuitous. The length
              in codons is given for each CDS.
              Usually the highest scoring match found by fasta -o is given for
              CDS which show significant similarity to other CDS in the database.
              The position of possible ribosome binding site sequences are given
              where these have been used to deduce the initiation codon. Gene
              prediction is based on positional base preference in codons using a
              specially developed Hidden Markov Model (Krogh et al., Nucleic
              Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
              of Bibb et al., Gene 30:157-66(1984) as implemented at
              http://www.nih.gov.jp/
              jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the
              correct initiation codon. Where possible we choose an initiation
              codon (atg, gtg, ttg or (att)) which is preceded by an upstream
              ribosome binding site sequence (optimally 5-13bp before the
              initiation codon). If this cannot be identified we choose the most
              upstream initiation codon.
              IMPORTANT: This sequence MAY NOT be the entire insert of the
              sequenced clone. It may be shorter because we only sequence
              overlapping sections once, or longer, because we arrange for a
              small overlap between neighbouring submissions. Cosmid 5f8 lies on
              genomic restriction fragment AseI-A bordered by cosmids 5H1 and
              4G10.

FEATURES             Location/Qualifiers
     source            1..33285
                       /organism="Streptomyces coelicolor A3(2)"
                       /strain="A3(2)"
                       /db_xref="taxon:100226"
                       /clone="cosmid 5f8"
     misc_feature      complement(2..109)
                       /note="nominal overlap with cosmid SC5H1 between bases
                       1..108."
     gene              complement(493..1452)
                       /gene="SC5F8.02c"
                       /db_xref="SC07292"
                       /note="SC07292"
     CDS               complement(493..1452)
                       /gene="SC5F8.02c"
                       /note="SC5F8.02c"
                       aa. Similar to many including: Escherichia coli
                       SW:THD2 ECOLI(EMBL:M21312) threonine dehydratase catabolic
                       (EC 4.2.1.16) TdCB (329 aa), fasta scores opt: 609
                       z-score: 676.6 E(): 3.2e-30 39.4% identity in 312 aa
                       overlap. Contains a prosite hit to Serine/threonine
                       dehydratases pyridoxal-phosphate attachment site and a
                       Pfam match to entry PF00291 S.T. dehydratase,
                       Pyridoxal-phosphate dependent enzyme."
                       /codon_start=1
                       /transl_table=11

```

```

/product="putative threonine dehydratase"
/protein_id="CA893730.1"
/db_xref="GI:8347025"
/db_xref="SPTREMBL:Q9K4M2"
/translacion="MIAITEAARAAGVAVSGNHGIAVAVMAALDVKATVVMPTA
ORTGSPKARGATAKLLSLTEAARAAGVAVSGNHGIAVAVMAALDVKATVVMPTA
PARSVIEAEGALVRLTDGMSAFALVRLREGLTVHPDDPVPVAGQGTGVEF
ADADSLTDLVLSIGGGGLIAGVAAALRALRPGVRVGVETEGAEAMSRALAGPLT
VPLSSVVTLTSLAPSRLYLDHVAELVTEVLVVPDREAVRGSLALAHAKVWTEPAAG
CLLPARRVVERVGDAIGLVVCGGNATVGDMAVWADRFLGR"
complement(547..1413)
/gene="SC5F8.02c"
/note="Pfam match to entry PF00291 S.T. dehydratase,
Pyridoxal-phosphate dependent enzyme, score 210.10,
E-value 3.3e-59"
complement(1288..1329)
/gene="SC5F8.02c"
/note="PS00165 Serine/threonine dehydratases
pyridoxal-phosphate attachment site"
complement(1478..2155)
/gene="SC5F8.03c"
/note="SC07293"
complement(1478..2155)
/gene="SC5F8.03c"
/note="SC5F8.03c"
Similar to several proteins of undefined function
including: Deinococcus radiodurans TR:Q9RWP1
(EMBL:AE001920) conserved hypothetical protein (254 aa),
fasta scores opt: 467 z-score: 555.6 E(): 1.8e-23 36.6%
identity in 205 aa overlap and Streptomyces coelicolor
TR:CA870638 (EMBL:AL137242) SC8F4.12C (239 aa), fasta
scores opt: 397 z-score: 474.2 E(): 6.1e-19 36.1% identity
in 183 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CA893731.1"
/db_xref="GI:8347026"
/db_xref="SPTREMBL:Q9K4M1"
/translacion="MKGDLFSSEHMVQPPATPMVTENSKCIYAVNGEMLARQAGMI
AVRGNLQFERKQGVGKMLKRAVTEGGLPLMVRGQGEAFANEAQCFVEVEPDGE
FTVNGRNLVCFDASLRIATYKVGSIAGGLFNSVFTQGRGLVCEGNPLVIVSQ
QYFVHYDTDAVVGWSAGLTSLSRQSGISGMLRGSGEAVQLVQEGFVVRPSEAT
PQKPOQH"
complement(2167..3471)
/gene="SC5F8.04c"
/note="SC07294"
complement(2167..3471)
/gene="SC5F8.04c"
/note="SC5F8.04c"
Similar to Streptomyces coelicolor TR:Q9XW99
(EMBL:AL096743) putative transport associated protein,
SCI7.21 (399 aa), fasta scores opt: 428 z-score: 449.2
E(): 1.5e-17 38.3% identity in 209 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CA893732.1"
/db_xref="GI:8347027"
/db_xref="SPTREMBL:Q9K4M0"
/translacion="MSDLVPGNVPLPGGVSVRVPGFVDSALVTDGSKVGGDADF
VFNQPEAPGARLDDTLVDPARLRGARVTVAVGSPDCTPLGALPSPTVLVTD
RGTVARTFPAFRGRTVLLVLFYRGGKVKVRLGQGVADGLAGLARDFGVEVDD
APPEPATAPEPLTPGADSDGFLNLSNARAAGSPARVDFPRLRSARAHAAVADG
TSLIETRDGVSQHVQVVSFAVLYTGEHLVSGPTPAEFVAYCLRAERTRLTHTDA
FTHAGWACTVGGSGDTYTWALMAVLPDGLARTTAEVVGLTNRAGAGLALAD
ARLTAQAQHSADMTVTRDFVSHTDPDGKPFWDRAAAGADRRSVGENIACQGRSPAEV
CEWMNSPGRHAILAGETHIGVGLAGGRAGTYWTQLGG"
complement(3563..6529)
/gene="SC5F8.05c"
/note="SC07295"
complement(3563..6529)
/gene="SC5F8.05c"

```

/note="SCSF8.05c, possible LuxR-family regulator, len: 988
aa. Similar to several other putative regulators from
Streptomyces coelicolor including: TR:Q9XAN4
(EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305
z-score: 338.7 E(1): 7.7e-11 26.4% identity in 910 aa
overlap and TR:Q92573 (EMBL:AL035569) SC8D9.18 (1091 aa),
fasta scores opt: 857 z-score: 926.1 E(1): 0.36.5% identity
in 1077 aa overlap. Contains a Prosite hit to P80017
ATP/GTP-binding site motif A (P-loop) pfam match to entry
PF00196 GerE, Bacterial regulatory proteins, luxR family."
/codon start=1
/transl_table=11
/product="putative LuxR-family regulator."
/protein_id="CAB93733.1"
/db_xref="GI:8347028"
/db_xref="SPTREMBL:Q9K419"
/translation="MGVGPOTSLPVGREDELARLTGVLERAGARAVLVAGDA
GVGKRTLHEAGRAAAGATTVLGHCVLDGVDGLPYLFTTEILGVLAADRFPAVLA
GHPVADRLUGGGDDTDAPSRLLRPFVVAALLTELADVAPLLVLEDLRWQDSR
DLLRFLGRVLQRPAGGPHRLALFASRADLHRRPLRLPLLAELVRLPGVERLE
LRPLPDSOVARLRLRRRLPDSTVRIVERAEGNAFYAEELVAATDAPAGVPSGL
ADVLRIEQLSETAQOVLRTAAVAGRRVGHDLRDVGLPEEELSALRAVERQLL
VSGDGYSPFHARAEAVADLLPGEARLHGAFARLLAGPDRSASARAHYRE
SHDLPALASLEAADHQRVAGPABEELRHVERAULDLMTAVDAARAGPDAVTLTR
ASAAVHAGEHLRAVSLRSLAGLGQDADLEAARVRYTLAGNLLSVNLESAYIS
SEALGLI PAEPSPSTVWAAATHVTAAQVGENETALRVARRALAEELAVTDARAD
LLISLTLEGNNSTPEGRLELARELARAGNAPVELRALFNLAI GCPSPGDLLEC
LPWASGDLRRAGSLSSPYPREMYRLRLVQVTLGHWDVLESAEHAGERSAVDG
HALGFLVARRAGDSVADRARALLGPPDPWARMVAGVVLTDAAALRGDAEADRW
MRSSVETITEAGRTPTVTLRLATLALSADVTTELRRAGDAGVARWSDTAELLAD
ARWAGHGDIMPQFEGQAWLAEAEWSVAGAAPDPAAWRAVAATFTGDTYERA
RCRLRLADALLAAREEATEAGVREARDLGLATVLRERVDLVRGRLAGPARAG
AGAAVLTAREQDVLRLALGHSNRRIIGELFISAKTASVHVSNI LAKLNAASRTAEVA
VAHRQGLVAPEPTSSH"

Query Match 9.2%; Score 65.6; DB 1; Length 33285;
Best Local Similarity 55.7%; Pred. No. 1.3e-05;
Matches 147; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

QY 432 GCCCATGTTGATGGCAATCGTCTGACCTGTTGACCCGATCGTATTGGACCGGTT 491
Db 18542 GCCCGTGTCTACCTGCTGGCGCTGACCTGCTCAACCGCACCGCTTACCTCGACACCGT 18601
QY 492 TGTGTTTATCGCGCGCGTCCGCGCGCAATACCGCGCACCGGACGCTGATTTTCGCGCG 551
Db 18602 GTTCTGCTGGGCTCGTCCGCGCGC---CGACCGGGCGCGCTGCGCTTGGCCT 18658
QY 552 TGGCGGTTGCGCGCAAGCTGATCTGTTCCCGCTGTTGGGTTTTCGCGCGACGACGATT 611
Db 18659 CGGAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 18718
QY 612 GTTACGCGCGCTGCTGCGCGCAAGCTGTTGGGCTGATCAAGCTGCTGCGCGCGCTGCT 671
Db 18719 CGGCGGCTTCTGTTCCGCGCGCGCTGCTGCGCGGCTGCTGCGCGGCTGCTGCGCGCAC 18778
QY 672 GATGACCGCATTTGGCCATCAAACT 695
Db 18779 CATGATCGTCTCGCGGCTCCCT 18802

RESULT 11
AE004852
LOCUS
DEFINITION
Pseudomonas aeruginosa PA01, linear BCT 30-AUG-2000
genome.
ACCESSION
AE004852
VERSION
AE004852.1
KEYWORDS
GI:9950587
SOURCE
Pseudomonas aeruginosa.
ORGANISM
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 10542)

AUTHORS

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Coulter, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,
Brody, L.D., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)

TITLE

20437337

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 10542)
Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Coulter, L., Tolentino, E.,
Westbrook-Wadman, S., Yuan, Y., Brody, L.D., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Salter, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission

TITLE

Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA

FEATURES

source

Location/Qualifiers

1..10542

/organism="Pseudomonas aeruginosa"

/strain="PA01"

/db_xref="taxon:287"

/complement (139..612)

/gene="PA4360"

/complement (139..612)

/gene="PA4360"

/codon start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG07748.1"

/db_xref="GI:9950588"

/translation="MRLOVQLQELKGSQROADVSNQFGGKRLDOLASDLKAQO
ESAQVLAQLDGKLOTLAAOEKALQVELKTNELKALGDVAALKQGNPSQTI
KGTQDMLVLRSQLDNRPAGGNTAEFDAPRTQVMRNINTLQAVQVQLQQLNTR"

/gene="PA4361"

/complement (1529..2518)

/gene="PA4361"

/codon start=1

/transl_table=11

/product="probable oxidoreductase"

/protein_id="AAG07749.1"

/db_xref="GI:9950589"

/translation="MRILVTGATGFTGGRFARFALQGLSVRSVGRADAVEHLVARG
AEFVPGDLADPALVLRLCDVAVHACAGVGVGPRFLAANGLASVVEACMRQ
KYRRLVHLSPSIYFDGRDLDLNEEYVPRPSDHYGATKYQAEQLVLSARDLGLVL
ALRPFVVGAGDTSIFPRMIQAHKGRKLILGNLRVDFTSVHNLDAFSCLLAGE
PALGVYINISGQVPFVDVAVNYVMRQLDLPVGHLPYAVGVYGLAALNEGVCRLPG
RPEVILFRILGMVAAMKFTLDINRAREYLDYDPRVSLMTALDEFCAWMRAQGI"

2703..3761

/gene="PA4362"

2703..3761

/gene="PA4362"

/codon start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAG07750.1"

/db_xref="GI:9950590"

/translation="WASHMWIYGANGYTRLVAEQARQREGLTFLGGRNPAALHAIGS
QIGLEQRFVLDGDPQACREALDQKVVAHACGPFATSTPMIAACRAATYVDITGE
IAVFEQAHAGDAEAREAGTVCGVDFVPTDCLAACIKALPAQRLALGFDGSG
LSTGAKTSVEGLKFGKIRENRLRDVPLGKRRIDIDFGRLRHAVTIPMGDVATAY
YSTGIPDIEVLPAPELLALGRLIDPLPLGRQVQDWLKGQVDKRIAGPDQARE
RLRTVWGEARNRERRTARLETANVYDITLHGVLAVRHLLDYQGGPGGYTPSRL
GACVESLPGSGRTIVIG"

complement (3762..4664)

/gene="iciA"

/note="PA4363"

gene


```
CDS      complement(3762..4664)
        /gene="icia"
        /codon_start=1
        /transl_table=11
        /product="inhibitor of chromosome initiation IciA"
        /protein_id="AAG07751.1"
        /db_xref="GI:9950591"
        /translation="MILFDYKLLAALAAVVEGFFRAAQAQGLSQSAVSQRIKLEA
RVGQPLVRETPPHPTDLGRLLNHVQVRLLEGDLQRYWPNLDEGAPRLIALNA
DSLATWAAVAGDFCAERRVLLDUVEDQVGLKRMAGVAGCVCGSARPVAGARSL
LLGAMRVGLGDDPFIDARHFRGVAAALAGVPAIVFGPDLLQHRFLKDLGVGGFI
HHLCSSEGFRVRLTAGLWGLVPERQVQGLARGELVLLPGQVIDIVPLYWYWRNG
GELLASITHELLARAGDGLVRVSG"
        4762..5163
        /gene="PA4364"
        4762..5163
        /gene="PA4364"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAG07752.1"
        /db_xref="GI:9950592"
        /translation="MAGETSLRTLLRTLPENPGDYVFCACAGALPNGAEPLASFR
ERGLTVLBERQARLGLPYEYVAAMITLVHSSLTAVGLTAFAFALAEAGISCNV
MAGYFHDHLFVARDEGRRLAALVQLAAEAH"
        5165..5767
        /gene="PA4365"
        5165..5767
        /gene="PA4365"
        /codon_start=1
        /transl_table=11
        /product="probable transporter"
        /protein_id="AAG07753.1"
        /db_xref="GI:9950593"
        /translation="MWSYLNGLILVAAGLIIAIGAQNAFVLAQSLRREHLSVAALCV
FCDVAVSLGVFGFLKLLLENPTLIIARWGGAFLTWYGLKALLRALRPDALGNRAE
TCPRSKAVGLAALAVLLNPHVLDVTLVIGSLGAQQAAPGAYALGAASASLWVFA
LALCAAWLAPLARPATWRLDLLVAAWMLGMAQLLFRG"
        6114..6695
        /gene="sodB"
        /note="PA4366"
        6114..6695
        /gene="sodB"
        /codon_start=1
        /transl_table=11
        /product="superoxide dismutase"
        /protein_id="AAG07754.1"
        /db_xref="GI:9950594"
        /translation="MAPELPLPYEKNALPHISAETLYEYHDKHNTYVYVNLNLP
GTEPEGKSLERIVKSSGIFPNAAOVWNHTFYWNCISPNGGQOPTGALADAINAARG
SFDKFEFTKTSVGTSGCGWGLVKKADGSLALASTIGAGNPLTSGDTPLLTCDWVE
HAYIDYRNLRPKYVEAFWNLVWDFVAKPAA"
        6876..8939
        /gene="PA4367"
        6876..8939
        /gene="PA4367"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein"
        /protein_id="AAG07755.1"
        /db_xref="GI:9950595"
        /translation="MKLSYRHSLSLKLRLVLLAALAVGVLSQAQIVPDYAKAKQAV
SSDAQRIILAVRDPSTQVSYSDREMAQVLEGLFOHEAVRQASIGHGPEMLAEKR
PLLLPLFWLTDPLILQBERTPISRLICRPYSEYIGDKLTLDTPAGNFVFTSEII
FISGLRLAAGLVLFVYHMLTKPLSKIEHLVSINPDRPSQHPQLKGHNEL
GLWTTANQALASIESHSGRREAEENLRIISQYDFLTGPNQLQQDLQIDLGAG
RQQRVAVLCGLDDFGINEQYTYQLGQDQLIALADRLGRHSARLGLSGLDQDPA
LVQADIEQPYEAELAISILDGLEAPDEIDOEVRLEATIGITLFPEDGETKLLQK
AEQWTLAKTRSRNRYOFYIASVDSEMRRELEKLDALQHEHLVYQPOVDYED
HRVVGVALLRWQHLHGFVPPDLFIPILAQNGSIFSIGEWLDQACRQLREHWDQF
DILRMVNLSTVQLHNALPRVSNLLOVTRKLPARSLEVTETGLMEDISTAAQHLL
SLRAGALIAIDDFGTGYSSLSYLSKLSPLDKIKDKSFVQDLQDEODATIVRAIQIOL
KSLGMQVIAEGVETAQEAYIIAEGCNEGQGYLSKPLPARELTQYVLRKQARLSQAT
```

```
SSERP"
complement(8940..9842)
/gene="PA4388"
complement(8940..9842)
/gene="PA4368"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07756.1"
/db_xref="GI:9950596"
/translation="WDFDLWLFIAASGLCAWAGQWQFVHRLARALLGDVPTSRIS
AQGFVELYGLVEEGEPGLQAPLTGKPLWNRVRIEVEKESGREKAWRTVDKGLYSE
SPESLRTADACLVDPRAEVRPLTOQRWEAFREAPIDTLRLLESFGTLVGGELVRY
TEERLHAGELPYALGEPFRSSGGROGLDARAGAVIREWKGDPHGLLARPDSDNGGE
LDERWNRVRLAARLEADNRHASSAAPTNRHRLKPGEAHPFILLSQGGDDLLAVQURR
QAAGAALLCIGSALLAGWLFSLRF"
complement(9844..10416)
/gene="PA4369"
complement(9844..10416)
/gene="PA4369"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07757.1"
/db_xref="GI:9950597"
/translation="MSLTAIAFWVLLLAGYAVLVNGLVRLKHAVGKAWANIDVLL
KQRHDELPLKVEACKQYMOHERATLERTVIAARNAVASARERADYNALGKAESGLRAGL
GQLFALAENYPOLKANESFQILSRISGLENGIADRRELYNEAVNLNNVRIEQFPDLL
IARVAGFKAAELLOFSEAEKADVDLKALFG"
BASE COUNT      1755 a 3724 c 3387 g 1676 t
ORIGIN

Query Match      8.9%; Score 63.4; DB 1; Length 10542;
Best Local Similarity 55.0%; Pred. No. 4e-05;
Matches 149; Conservative 0; Mismatches 116; Indels 6; Gaps 1;

QY 440 TGATGGCAATCGTGTGACCTGTTGAACCGAATGGTATTTGGACGCGTTTGTGTTTA 499
Db 5502 TGGCGGCACCTGGCGGTACCTGCTCAACCCCACTCTATCTGATACCGTACTTCTTA 5561
QY 500 TCGCGCGCGTTCGGCGCAATACGGCGACACCGGACCGTGGATTTTCGCGCTGGCGCGT 559
Db 5562 TCGGTTCGCTCGCGCGCCACGAGG-----CGCGCGCGCGCTATGCTCGCGCGG 5615
QY 560 TCGCGCGAACCTGATCTGTTCCGCTGTGGTTCGGCGGACGACGATGTCAGCC 619
Db 5616 CCAGCGCTCGCTGATGTGTTCTTCGCCCTCGCGCGCGCATGGCTGCGCCCT 5675
QY 620 CGCTGTTCAGCCCCCAAGGTGTGGCGCTGATCAACGTCGTCGTGGCAGTTGTGATGACCG 679
Db 5676 GGTTCGGCGCGCGCGCCACCTGCGGCTGCTCGACCTGATGTTGGCGGCATGATGCTGG 5735
QY 680 CATTGGCCATCAAACTGATGTTGATGGGTTA 710
Db 5736 GCATGGCGCGCAACTGCTGTTCGGGGATA 5766

RESULT 12
AE008841/c
LOCUS      AE008841      24336 bp      DNA      linear      BCT 31-JUL-2002
DEFINITION      Salmonella typhimurium LT2, section 145 of 220 of the complete genome.
ACCESSION      AE008841      AE006468
VERSION      AE008841.1      GI:16421612
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2.
Bacteria; Enterobacteriaceae; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 24336)
McCllland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
```


Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R. K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

JOURNAL Nature 413 (6858), 852-856 (2001)

MEDLINE 21534948

PUBMED 11677609

REFERENCE 2 (bases 1 to 24336)

AUTHORS The *Salmonella typhimurium* Genome Sequencing Project.

TITLE Direct Submision

JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest

Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES

source
1. .24336
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
171. .733
/gene="ygfa"
/note="STM3061"
171. .176
/gene="ygfa"
/note="putative RBS for ygfa; RegulonDB:STWS1H003104"
185. .733
/gene="ygfa"
/note="similar to E. coli putative ligase (AAC75949.1); Blastp hit to AAC75949.1 (182 aa), 86% identity in aa 1 - 182"
/codon_start=1
/transl_table=1
/product="putative ligase"
/protein_id="AAL21936.1"
/db_xref="GI:16421613"
/translation="MIKRRALTPQQRFQGOAAARMLSFPVVAHTVAVFLSFD
GELDTQPLIQMRAGRVLPVLPFSPGNLLFLHYPSQSLVTLNRLKIQEPLDVR
DVLPALAKDLVLPVAFDDEGQRLGNGGGFYDRTLQNWQHKILPVGAHDCQLVEK
LPVKEWDIPLPAVTPSKIWEW"
complement (1114. .2359)
/gene="serA"
/note="STM3062"
complement (1114. .2346)
/gene="serA"
/EC number="1.1.1.95"
/note="similar to E. coli D-3-phosphoglycerate
dehydrogenase (AAC75950.1); Blastp hit to AAC75950.1 (410
aa), 96% identity in aa 1 - 410"
/codon_start=1

/transl_table=11
/product="D-3-phosphoglycerate dehydrogenase"
/protein_id="AAL21937.1"
/db_xref="GI:16421614"
/translation="MAKVSLEKDKIKRFLIVEGVHQKALESLRAAGVTNIEYHKGALDA
EQLKASIRDAHFIFGLRSRHLTEEVINAAEKLVAICFCITGNQVDLNAARKGIFVF
NAFFSNTRVAELVIGELLILLRGPVEAKAHRGVNKLAAAGSFARGKGLGIIGYG
HIGTQGLIAESLGHVYFDYENKLPNGNATQVQSLSELLANMSDVVSLHVPNASTK
NMGAKEIALMKPGLINAAAGTVDIPALCDALATKHLAGAAIDVFTEPATNSDP
FTSPCEFDNVLTPIHGSTQEAQENIGLEVAGKLIKYSNKGSTLSAVNPFVSLPL
HGGRRLWHIENRPGVITALNQIFAEQGVNIAAQYLQTSARMGVYVIDEADGDVAEK
ALLAAIPGTINARLLY"
complement (2354. .2359)
/gene="serA"
/note="putative RBS for serA; RegulonDB:STWS1H003105"
complement (2612. .3286)
/gene="rpiA"
/note="STM3063"
complement (2612. .3271)
/gene="rpiA"
/EC number="5.3.1.6"
/note="similar to E. coli ribosephosphate isomerase,
constitutive (AAC75951.1); Blastp hit to AAC75951.1 (219
aa), 98% identity in aa 1 - 219"
/codon_start=1
/transl_table=11
/product="constitutive ribosephosphate isomerase"
/protein_id="AAL21938.1"
/db_xref="GI:16421615"
/translation="MTQDELKKAAGVAAALQYVQPGTIVGVGTSTAAHFIDALGTWKG
QIEGAVSSDSASTKGLGIHFDLNEVSLGIYVDGADEINGHMVKIKGGGAALKG
EKIASVASKFICIAADKQVDILGKFLPVEVIMPARSAVARQIVKLGRPEYRONV
VTDNGNVILDVYGMELDIPALENAINAIPGVTVGLFANRGADVALIGTDPGVKTV
K"
complement (3281. .3286)
/gene="rpiA"
/note="putative RBS for rpiA; RegulonDB:STWS1H003106"
3407. .4318
/gene="iciA"
/note="STM3064"
3407. .3412
/gene="iciA"
/note="putative RBS for iciA; RegulonDB:STWS1H003107"
3425. .4318
/gene="iciA"
/note="lysR family; similar to E. coli replication
initiation inhibitor, binds to 13-mers at oric
(AAC75953.1); Blastp hit to AAC75953.1 (297 aa), 97%
identity in aa 1 - 297; transcriptional regulator of dnaA
and argK"
/codon_start=1
/transl_table=11
/product="inhibitor of replication initiation"
/protein_id="AAL21939.1"
/db_xref="GI:16421616"
/translation="MKRPDYRTLQALDAVIREGPERAAKLCITQSAVSQRIKOLEN
MGQPLLVNTVPPRTEQKGLALLRQVELLEEWLGDEQTSGLLUSLAVNADSL
ATWLPALAPVLAADSPIRLNQVDETRQLRERGVGVGAVSIQHALPCLVDKLG
ALDYLFVASKPFAEYFPNGVRSLLKAPAFDHLDMHQAFQONFDLPFGVPC
HIVNSEAQVQLARQCTTCMTPHLQIEKESLGLINLTGQLQRRLMYWHRFAPE
RMWRKYTDALLEYGHKVLQD"
complement (4584. .5341)
/gene="yggE"
/note="STM3065"
complement (4584. .5330)
/gene="yggE"
/note="similar to E. coli putative actin (AAC75959.1);
Blastp hit to AAC75959.1 (246 aa), 88% identity in aa 1 -
245"
/codon_start=1
/transl_table=11
/product="putative periplasmic immunogenic protein"
/protein_id="AAL21940.1"

```

/db_xref="GI:16421617"
/translation="MKFKVMAALAVGLSAMSQAQSELPEGPPIVITSGTASVDAVPDI
ATLAEVNAKDAATKQADERVAQYLPLEQNIKXDI SAANLRTQPDYDQNG
KSLIKGVRAVTVTRQLDKLNSLDGALKAGLINEIRSVLSGVAQPDYKOKARKA
AIDDAIHOAOALAGFNSKLGPSVSVRYHSYNYQSPSPVVRMKAADAAPVSAQETVEQ
PTIQPDQDVVFOLEPGTGTSTTAASTQ"
complement(5336..5341)
RBS
/gene="yggB"
/notes="putative RBS for yggB; RegulonDB:STWS1H003108"
gene
complement(5424..6072)
/gene="yggA"
/notes="yggA"
CDS
complement(5424..6059)
/gene="yggA"
/notes="similar to E. coli orf, hypothetical protein
(AAC75960.1); Blastp hit to AAC75960.1 (211 aa), 90%
identity in aa 1 - 211"
/transl_table=1
/product="putative LYSE family amino acid transport
protein"
/db_xref="GI:16421618"
/translation="WIGVYFOGFAALGAMILPLGPONAFVWNOGIRROYHLMIALICA
LSDVLISAGIFEGSALLMOSPLLALVTWGGVAFLLWYFGALKATAMSSNLELASAE
VMKGRKLIATMLAVTWLPHVYLDIFVILGSLGGLAMEPKRWFAFGTISAFSLWF
FGLALLAAWAPLRTAKAQRINILVGVWMLIAFQLAREVAHMAHLPN"
complement(6067..6072)
RBS
/gene="yggA"
/notes="putative RBS for yggA; RegulonDB:STWS1H003109"
gene
complement(6259..7131)
/gene="yggB"
/notes="yggB"
CDS
complement(6259..7119)
/gene="yggB"
/notes="involved in stability of Macs mechanosensitive
channel; similar to E. coli putative transport protein
(AAC75961.1); Blastp hit to AAC75961.1 (286 aa), 91%
identity in aa 1 - 285"
/codon_start=1

```

```

Query Match      8.8%; Score 62.8; DB 1; Length 24336;
Best Local Similarity 51.8%; Pred. No. 6.4e-05;
Matches 142; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 23 TTACAGGTCTGTTTGGGGCCAGTCTTTTACTCTCATCGGACCGCAGATGACTGG 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6043 TTCAAGGGTTTGGCCCTGGCGCGCATGATCTCTTCCGCTTGGCCGCAAAATGCGTTG 5984

Qy 83 TGATTAAACAGGATTAAGCGCGAGGACATCTGCGTCTTCTCGTGTGTTTAAATT 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5983 TGATGAATCAGGAATTCGTGCCAGTACCACCTCATGATACGCTGCTTTCGCGGTTAA 5924

Qy 143 CTGACGCTCTTTTGTTCATCGCGGCACTTGGCGGTGTGATCTTTTGCCATGCGCGGC 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5923 GTGATTTAGTCTGATTAGCGCGGTATTTTGGCGGTAGCGGTGCTGATGCACTGCG 5864

Qy 203 CGATCGTCTCGATTTATGCGCTGGGTGGCATCGCTTACTCTGTTATGTTTGGCGTCA 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5863 CGTGGCTCTGCGCTGCTGATCGATCGGCGGGTCCGCTTCTGTTATGATGACGTTTGG 5804

Qy 263 TGGCAGCGAAGACGCCATGACCAACAGGTGGA 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5803 GCGGCTGAAACGCGCAATGACGATATCTGGA 5770

```

```

RESULT 13
AE005522/c      10701 bp      DNA      linear      BCT 21-MAR-2001
LOCUS           Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 141
DEFINITION      of 290.
ACCESSION       AE005522 AE005174
VERSION         AE005522.1 GI:12517448

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
Source
gene
CDS
gene
CDS

```

Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 10701)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimallanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
11206551
2 (bases 1 to 10701)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.J., Davis,N.W., Lim,A., Dimallanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..10701
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/ecotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
133..2277
/gene="sbm"
/note="24254"
133..2277
/gene="sbm"
/function="enzyme; Degradation of small molecules: Fatty
acids"
/note="Residues 1 to 714 of 714 are 99.15 pct identical to
residues 1 to 714 of 714 from Escherichia coli K-12 Strain
MG1655: B2917"
/codon_start=1
/transl_table=1
/product="methylmalonyl-CoA mutase (MCM)"
/protein_id="AAG58043.1"
/db_xref="GI:12517449"
/translation="MSNEQEQMOQLANKELSRREKTVDSLAVQQTAEGLIKPLYTEADL
DNLEVTGLPGLPPVYRGPRATMYTAQPIITQYAGFTAKESNAFYRNLAAGQGL
SVAFDLATHRGYDSNPRVAGDVGTGVAIDTVDMKVLFDQIPLDKMSVTMNGAV
LPLVAFYITAEQGVTPDKLTGTQNDILKEYLCRNTYIYPPKPSMTIADIAWCS
GNMPRENTISISGYHMGAGANCVQVATADGIEYIKAAISAGIKIDDFAPRLSFF
FGIMDLFMNAMLRAARYLWSEAVSGFGAQPDKSLARTHTQCTGWSLSETPQPPYNNV
IRTTLEALATLGGTOSLHTNAFDEALGPTDFSAIARTQIIOESELCTVDPL
AGSYIESLTDQIVKOARAIIOQIDEGAMAKAIEAGLPKRMIEEASAREQSLDQK
RVTGVNKKYLDHEDTVDLEIDNVVRNEQIASLERIRATRDAAATAALNALTHAA
QHNEMLAAAVNAARVATLGEISDALEAFDRYLVPSQVTVGVIQAQSHQSEKSAFE
FOAIVAQTEQFLADNRPRRLITAKMLGDHDKAKVIASAYSDLGDFVDLSPMFSTP
BEIARLAVENDVHVVGASSLAAGHTLIPELVEALKKWKREDICVVGAGVIPPQDYAF
LOERGVAIYGGPTMLDSVRDVLNLSQHHD"
2270..3265
/gene="ygfD"
/note="24255"
2270..3265
/gene="ygfD"
/function="phenotype; Not classified"
/note="Residues 1 to 331 of 331 are 99.69 pct identical to
residues 1 to 331 of 331 from Escherichia coli K-12 Strain
MG1655: B2918"
/codon_start=1
/transl_table=1
/product="putative nucleotide-binding protein"

```

/protein_id="AAG58044.1"
/db_xref="GI:12517450"
/translation="MINEATLAESTRLRQGEHTLAQAMTLVESRHRHQALSTQLL
DAIMPCYCNLTGLVGTGPGAGKSTFLFAGMLLIRGLKVAIVADFPSPVGTGSL
GKTRNDLRARAAIRPVPSSGHLGASQARREMLLCEAAGDYDVIVETVGVQS
ETEVARNDLCFISQIAGGDDLLQKKGMEVADLIIVNKDGDNDNTNVAIRHMYE
SALHLIRKIDWQPRVLTCSALEKRGIDEIWHALIDFKTALTASGRLLQVROQOSVE
WLKQTEBEVLNHLFANEDFDYRYRQTLAVKNNTLSPTGLRLQSLSEFIQTQYFD"
gene
/gene="ygfG"
/notes="ygfG"
/notes="putative enzyme; Not classified"
/notes="Residues 1 to 275 of 275 are 100.00 pct identical
to residues 1 to 275 of 275 from Escherichia coli K-12
Strain MG1655: B2919"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAG58045.1"
/db_xref="GI:12517451"
/translation="NLSRRNLIKGFMSYQVNVVNTINKVAVIEFNYRKLNALSK
VFIDMLQALSDLNPEIRCIILRAPSGKVSFAGHDIEHLPSSGRDPLSYDDPLRQI
TRMIQKFKPIISMVGGVWGAFAEMSSDLIIAASSTFSMTPVNLGVYNLVGHI
NLTRDAGPHIVKELIFTASPIAORALAVGLNVVEVELEDFTLOWAHHISEKAPL
ALAVIKELRLVGEAHTWNSDEFERIQGMRRVDSYQSGMNAFLKRPNFVGH"
4085..5563
/gene="ygfH"
/notes="ygfH"
/notes="putative enzyme; Not classified"
/notes="Residues 1 to 492 of 492 are 99.39 pct identical to
residues 1 to 492 of 492 from Escherichia coli K-12 Strain
MG1655: B2920"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAG58046.1"
/db_xref="GI:12517452"
/translation="METQWTRTADAEAAIIQHNDVAFSGFTPAGSPKALPTAIARR
ANEQHEAKKQIIRLLTGASISAAADDVLSADAVSWRAPYQTSGLRKKINQAVSF
VDLHLSVAQMVNYVFGDIDVAVIEASALAPDGRVMTSGIGNAPTLLRAKVIIE
LNHYDPRVLAELADIVIPAPPRNSVSIHAMDRVGYVQIDPKTIVAVFENLDP
AGMLDKQNPQCQIADNVVFLLEQMAHGRIPPEFLPQSGVGNINNAVWARGENP
EIPPMWTSVLQESVSHLETKIGKISASASLSLTISADSLRKIYDNDMDYFASRIVLRP
QEISNNPBIIRLGVIALNVGLEFDIYGHANSTHVAGVDLMNGIGSGDFERNAYLSI
FMASIAKEGKISTVPMCSHVHDNEHSVKVIIIEQGIADLRGLSLPQARTIIDNCA
HPMYRDYLHRYLENAPGGHIIHDLISHVDFLHRLNIATGSMGLG"
complement(5878..6456)
/gene="y4258"
/complement(5878..6456)
/gene="y4258"
/notes="y4258"
/function="putative regulator; Not classified"
/notes="Residues 1 to 192 of 192 are 99.47 pct identical to
residues 6 to 197 of 303 from Escherichia coli K-12 Strain
MG1655: B2921"
/codon_start=1
/transl_table=11
/product="partial putative transcriptional regulator
LYSR-type"
/protein_id="AAG58047.1"
/db_xref="GI:12517453"
/translation="MDIFISKMRNFILLAQTNNAIAEAKIHWASPPFGKSIALAE
QIGTLFTRKNDQISLAKAGELYKQLFPVYQRLSAIDNEIHNCSRSREIIVGIDNT
YPTIFDOLISLQKYGVTAPQVPEFSENGVIDNLFDRQLDFIISPQHSARVQELN
LTISELPLRLNLFVSRREYERQELLQELP"
complement(6623..7363)
/gene="yggE"
/notes="yggE"
/notes="putative enzyme; Not classified"
/notes="Residues 1 to 246 of 246 are 100.00 pct identical
to residues 1 to 246 of 246 from Escherichia coli K-12
Strain MG1655: B2922"
/codon_start=1
/transl_table=11
/product="putative actin"
/protein_id="AAG58048.1"
/db_xref="GI:12517454"
/translation="MKFKVIALAALMGISMAQAQNELDPGPHIVTSGTASVDVDPDI
ATLAEVNVAAKDANTAKKQADQADQVAYQISFLELQIAKDISSANLRTQPDYDQD
KSLIKGYRAVTVTEVTLRLQDLKNSLLDGLAKGLNEIRSVSLGVQAQPDAYKDKAKKA
AIDNAIHQAQELANGFRLKGLPVYSRVYHSVYQPSFPMVMKADAAPVSAQTYEQEA
AIOFDQDVVFQLEPVDQOPAKTPAAQ"
complement(7456..8091)
/gene="yggA"
/notes="yggA"
/notes="putative enzyme; Not classified"
/notes="Residues 1 to 211 of 211 are 98.10 pct identical to
residues 1 to 211 of 211 from Escherichia coli K-12 Strain
MG1655: B2923"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58049.1"
/db_xref="GI:12517455"
/translation="MFSYYPQGLGALGANMLPLGPNQAFVFMNQIRQYHIMTALICA
ISDLVLTICAGIFGSGSPWLLALVTGQVFLVLTGFGAFKTMAGSNIELASAE
VLKQGRWKIATMLAVTLNPHVYLDTFVVLGSLGQDQVXPKRWALGTISASFLWF
FGLAILAAMLAPELRITAKSQRIINLVVGVCMWFIQLQARDGIAHAQALES"
complement(8230..9090)
/gene="yggB"
/notes="yggB"
/notes="putative enzyme; Not classified"
/notes="Residues 1 to 286 of 286 are 100.00 pct identical
to residues 1 to 286 of 286 from Escherichia coli K-12
Strain MG1655: B2924"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAG58050.1"
/db_xref="GI:12517456"
/translation="MEDLVNVDISINGAGSLVANQALLSYAVNVAALAIITVGLII
ARMSNVRLMISRKIDATVADFLSALVRYGIIAFTLIAALGRVGVQTASVAVLGA
AGLAVGLALQGLSLNLAAGVLLVMFRPFRAGEVVDLGGVAGTVLSVQIFSTTWRADG

```

Query Match 8.7%; Score 62; DB 1; Length 10701;
 Best Local Similarity 51.4%; Pred. No. 9.1e-05;

Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy	19	TTCAATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCCATCGGACCGCAGAAATGTA	78
Db	8079	TACTTTCAAGGTCTTGCTTTGGGGCAGCTATGATTCGCGCTCGGCCCCACAAAATGCT	8020
Qy	79	CTGCTGATTAAACAAGGAATTAAGCGCGAAGGAATCAATTCGCGGTTCTTCTCGTGTGTTTA	138
Db	8019	TTTGTGATGAATCAGGCATTCGTCGTGTCAGTACCACATATGATTCCTTACITTCGCT	7960
Qy	139	ATTTCTCAGCTCTTTTGTTCATCGCCGCGCACCTTGGCGGCTTGATCTTTTGTCCAATGCC	198
Db	7959	ATCAGCGATTTGGTCTCTGATTTGGCGGGGATTTTTCGTGTAGCGGCTTATTGATGCAG	7900
Qy	199	GCGCCGATCGTCTCGATATATGCGTGGGTGGCATCGCTACCTGTTATGTTGCTTGGC	258
Db	7899	TCGCCGCTGTTGCTGCGCGTGGTCACCTCGGGCGGCGTAGTCTTCTTGTGCTGATGGT	7840
Qy	259	GTCAATGCGCAGCAAGACGCGCATGACAAACAAGGTGGA	296
Db	7839	TTTGGCGCTTTTAAACAGCGATGATGATTAATATGGA	7802

```

RESULT 14
AP002563/c
LOCUS      26658 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 14/20.
ACCESSION AP002563 BA000007
VERSION   AP002563.1 GI:13363121
KEYWORDS
SOURCE    Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
          DNA.
ORGANISM  Escherichia coli O157:H7
REFERENCE Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
AUTHORS   Escherichia.
          1 (sites)
Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
Kurokawa,K., Ishii,K., Hattori,M., Tatsuono,I., Abe,H., Iida,T.,
Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
Sasakawa,C. and Shinagawa,H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
JOURNAL   20198780
MEDLINE   2 (sites)
REFERENCE Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
AUTHORS   Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
          Hayashi,T.
TITLE     Comparative analysis of the whole set of rRNA operons between an
          enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
          Escherichia coli K-12 strain MG1655
JOURNAL   20557356
MEDLINE   3 (sites)
REFERENCE Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
AUTHORS   Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
          Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
          Shinagawa,H.
TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
          Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
          O157:H7 strain derived from the Sakai outbreak
JOURNAL   Gene 258 (1-2), 127-139 (2000)
MEDLINE   20564182
REFERENCE 4 (sites)
AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
          Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
          Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
          Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
          Shinagawa,H.
TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
          O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL   DNA Res. 8 (1), 11-22 (2001)
MEDLINE   21156231
REFERENCE 5 (bases 1 to 266658)
AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
          Hayashi,T.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
MEDLINE   Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
REFERENCE Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
          URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
          Fax: 81-6-6879-2047)
          genome project.
FEATURES   Location/Qualifiers
           1..266658
            /organism="Escherichia coli O157:H7"
            /strain="O157:H7"
            /sub_strain="RIMD 0509952"
            /db_xref="taxon:8334"
            complement(244..693)
            /gene="EC83650"
            complement(244..693)
            /gene="EC83650"
gene
CDS

```

```

/notes="similar to B2790 ECOLI gi|1789154 percent identity
100 in 149 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37073.1"
/db_xref="GI:13363122"
/translation="MABIGIFVGTWYGNLLVAEAEAILTAQGHKATVFDEPELSDW
LPYQDKYLVVVTSTTQGGDLPSIVLFGIKDLSGLGFPNLRVGVIALGDSSVYVPCN
GKGQDALLQEASQAVRGVEMLLIDASENPEPETESNPWVEQMGTLIS"
complement(711..1493)
/gene="EC83651"
complement(711..1493)
/notes="similar to YQCB ECOLI gi|1789155 percent identity
99 in 260 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37074.1"
/db_xref="GI:13363123"
/translation="MLEILYQEWLVAVNKPQGLVHRSWLDREKVVVWQTVRDOIG
QHVPTAHLDRPTSGVLLMGLSSEAGRLAQOEQHOIKRVYHAIVRGWLMEAVLDY
PIVEELDKIADKFAREDKGPQAPVTHYRGLATVEMPVATGRYPTTRYGLVLEPKTGR
KHLRDLAHLRHPITIGDSKGLDKQNRGAHFGQLRMLHASQLSLTHPTTGEPLT
IHAGLDDTVMQALSQFGKGLLPENERVEFSAPSGQDGRSS"
complement(1493..1822)
/gene="EC83652"
complement(1493..1822)
/notes="similar to B2792 ECOLI gi|1789156 percent identity
98 in 109 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37075.1"
/db_xref="GI:13363124"
/translation="MTTHDRVRLQLQALEALLREHQHWNRDEFPQHFNSTOPPFMDT
MEPLEWLQWLIPRMHDLNNQPLPGAFAPVYEMALATDHPQRLILAELEKIDA
LPADDAS"
complement(1865..2233)
/gene="csrB"
complement(1865..2233)
/notes="inhibitor of CsrA"
/product="CsrB RNA"
complement(2444..2989)
/gene="EC83653"
complement(2444..2989)
/notes="similar to SYD ECOLI gi|1789157 percent identity 98
in 181 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Syd protein"
/protein_id="BAB37076.1"
/db_xref="GI:13363125"
/translation="MDDLTAQALKDFTARYCDAAWHEHKSPLSEELYGVSPSCIIST
TEDAVYVQPPFTGEQNVNAVERAFIVITQPTHTTFTTQPADGMAHQFGDKLTLQ
TWSDDFRVRQENLIGHLVTKRKLPLTLTATLEETELEVSVCNLSGEVCKETLGT
RKRTHLAPNLAELFNQLKPLL"
complement(3057..3905)
/gene="EC83654"
complement(3057..3905)
/notes="similar to YQCD ECOLI gi|1789158 percent identity
98 in 282 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental

```

```

/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37077.1"
/db_xref="GI:13363126"
/translation="MSSYANQALAGLTGKSTYRDYDASLIQGVPSRLNRPGLG
KADNLPFGDTIDWLYELUSWLNAGLPQVAVGHVELDYTVNLIESKSFKLYLNSFNG
TRFNWMDVROTLRDLSTCAQGEVSVALYRLDELQPGIHFNGTCTDDOITIDNY
BFTDYLENATSGEKVVEETLVSHLNSCLITHPQWMSIQIQRGQIDREKLRY
LVSPRHNEFHEQCVERIFNDLLRQCPQKLSVYARYTRRGGLDINPWSNSDFVPS
TRLVRQ"
4017..5381
/gene="EC83655"
CDS
4017..5381
/gene="EC83655"
/notes="similar to YGDH ECOLI gi|1789159 percent identity
99 in 454 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37078.1"
/db_xref="GI:13363127"
/translation="WITHISPLGSDMLSOLEVMDLKTASSDIYQLFRNCSLAVLNS
CSLTDNSKELSRFENDIVLRERRGVRKSLINPPEAFVDGRIIRALQANLPAVIR
DILFYGGIHTNVRFPNLDNSVHITNLVFSILRNARALHVGAPNNVCGGSHIN
ENBYLARRVGNQGLRELNICTGCGFAMEAPKGAAGVHAQYKDSRFIGNTPS
ITAAEPNPLVNELIIMPDIEKLEAFVRIAHGIIIPGGVGTAEELLYKILGMLNPA
NKDOVLPLILGPESADYPRVLDFVVTHTGENARRHYRIIDDAEVARQMKSNP
LVKRNRRDGDAYSFNMSRIADPLQMPFEPHENMANLKLPPQPEVLAADLRAP
SGIVAGNVKEIGIRAEIEFGYKINGDEIMRRMDLLQGFVAQHRMKLPGSAIYPC
BIC"
5938..7227
/gene="EC83656"
CDS
5938..7227
/gene="EC83656"
/notes="similar to SDAC ECOLI gi|1789160 percent identity
100 in 429 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="probable serine transporter"
/protein_id="BAB37079.1"
/db_xref="GI:13363128"
/translation="METTQTSTIASKDSRSARWRTDTMMLGLYGTAGVFLPLIN
AGVGMPLIIMALPMTFAHRGLTRFVLGKNPGEDITEVVEHFGIGAKGLIT
LLYFAIYPLLVSALTNTVESFMHQGMTPTPRALLSLILVGMWTVIRFGEOM
IVKMSILVFPFVGLMLALYIPQWGALETLSLDTASATNGLWMLTILWLAIPVM
VFSNHSPIISSYANHFNPAVIAWMAPIIAITKSFGLHGLARGEGFMWIK
DLAAKQNTSILSYANHFNPAVIAWMAPIIAITKSFGLHGLARGEGFMWIK
SLRGKSIKINKRITALPMLVTWIVATNPSILGMETLGGPIIAMILLFMPWY
AIQKVPAMKYSGHISNVFVVMGLIAISFYSLFS"
7285..8652
/gene="EC83657"
CDS
7285..8652
/gene="EC83657"
/notes="similar to SDAA ECOLI gi|1788116 percent identity
77 in 455 aa, similar to SDAB ECOLI gi|1789161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"
Query Match      8.7%; Score 62; DB 1; Length 266658;
Best Local Similarity 51.4%; Pred. No. 0.00014;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
QY 19 TCAATTACAGGTCCTTTTGGGGCCAGTCCTTTTACTGTCCATCGGACCGCAAGATGTA 78
Db 161581 TACTTCAAGGTCCTTGTCTTGGGGCAGCATGATGTCCTGCGCTCGGCCCAAAATGCT 161522
QY 79 CTGTGTGATTAAACAGGAATTAAGCGCAAGGACCTCATTTGGCGGTCTTCTCGTGTGTTA 138
Db 161521 TTTGTGATGATCAGGCAATTCGTCTGTCAGTACCAATTATGATTCCTTACTTTCGCT 161462
QY 139 ATTTCTGACGCTTTTGTGTCATCCCGCGCACCTTGGCGGCTTGATCTTTTGTCCAAATGCC 198

```

```

Db 161461 ATACGCGATTGGTCTCTGATTTCGCGGGGATTTTGGTGGTAGCGGCTTATTGATGCAG 161402
QY 199 GCGCGGATCGGCTCGATATTATCGCTGGGTGGGTGCGATCGCTTACCTGTTATGTTGGCC 258
Db 161401 TCGCGCGTGGTTCCTGGCGCTGGTCACCTGGGCGCGGTAGTCTTCTTGGCTGTGATGGT 161342
QY 259 GTCATGGCAGCGAAGACGCCATCACAAACAAGGTGGA 296
Db 161341 TTTGGCGCTTTTAAACAGCGATGAGTAGTAATATGCA 161304

RESULT 15
ASU65741
LOCUS
DEFINITION
ASU65741 4033 bp DNA linear BCT 11-SEP-1997
Aeromonas salmonicida orfC replication inhibitor (icia), N-acyl
homoserine lactone synthase (asaI), transcriptional activator
(asaR) and YggA (yggA) genes, complete cds, and 2'3'-cyclic
phosphodiesterase (cdpB) gene, partial cds.
ACCESSION
U65741
VERSION
U65741.1 GI:1519234
KEYWORDS
Aeromonas salmonicida.
SOURCE
Aeromonas salmonicida.
Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
Aeromonas.
REFERENCE
1 (bases 1 to 4033)
Swift, S., Karlyshev, A.V., Fish, L., Durant, E.L., Winson, M.K.,
Chhabra, S.R., Williams, P., Macintyre, S., and Stewart, G.S.,
Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
identification of the LuxRI homologs AhyRI and AsaRI and their
cognate N-acylhomoserine lactone signal molecules
J. Bacteriol. 179 (17), 5271-5281 (1997)
JOURNAL
MEDLINE
97431471
PUBMED
9286976
REFERENCE
2 (bases 1 to 4033)
Swift, S., Karlyshev, A.V., Fish, L., Durant, E.L., Winson, M.K.,
Williams, P., Macintyre, S., and Stewart, G.S.A.B.
Direct Submission
TITLE
Submitted (01-AUG-1996) ABFS, University of Nottingham, Sutton
Bonington Campus, Leicestershire LE12 5RD, England
FEATURES
Location/Qualifiers
1..4033
/orIGIN="Aeromonas salmonicida"
/strain="NCIMB 1102"
/db_xref="taxon:645"
/complement (30..929)
/gene="icia"
/complement (30..929)
/notes="icia; similar to Escherichia coli Icia"
/codon_start=1
/transl_table=11
/product="orfC replication inhibitor"
/protein_id="AAB70016.1"
/db_xref="GI:1519235"
/translation="MKALDYKLLALLDVAWQNFERRAQRHITQSASORIKOLEQ
QPAEPLIRSQPLQATPLQKLLAHYRQVRLQELAGIADDEQAPRVSIAVNAQ
SLATFPLAPALPELLEQHPLELNLVDDECTLDVRQEGAFGAVSLHGQAGCCVDE
LGEMRYLLTASPAFVARHPFAGLTPAALAKTAPAVDQDDMHVSMARHFGLEPGGY
PCHTVRSSEAFVMAEQGLAYCLTPELQIRQQAQGLILLDLSPSHLLIPLYWHRWVL
ERGLHQISQRLISGRRALQPG"
/complement (1195..1818)
/gene="asaI"
/complement (1195..1818)
/notes="asaI; similar to Photobacterium fischeri LuxI"
/codon_start=1
/transl_table=11
/product="N-acyl homoserine lactone synthase"
/protein_id="AAB70017.1"
/db_xref="GI:1519236"
/translation="MLVFKGLKHEPRWEVENELYRFRNRFSDRLGMDVESHRLGEQ
DSFDTPDTHWLVLEDEGLCGCIRLLSCAQDYMPLSIFPTALAGEAPRSSDVELTR

```

Job time : 2268.29 secs

LAI DAN RAPMNGVSELTCVIFREVVAFARAKGIRELVAVSLPVERIFRRLGLPIE
RLGHRQAVDLGAVGVGIRHILDERFARAVGHPMQGEYADARELVTE"

gene 1881..2663
/gene="asaar"
CDS 1881..2663
/gene="asaar"
/note="AsaR; similar to Photobacterium fischeri LuxR"
/codon_start=1
/transl_table=11
/product="transcriptional activator"
/protein_id="AAB70018.1"
/db_xref="GI:1519237"

/translation="MKQDQLLEYLHFTSVTGDRLAELIGRTGLMGDYDYRFALLI
PMSQRPVILFNQCPSPSWQVQYTNHMLACDPIIQARKQTLPIYWRDERARFLQ
EGSDVWGLAAEFLGRNGISFPLHGAAGENGILSFITAEASDDLLESPILSWMSN
YIFAAIRIVRVEDPQALDRETECLFWASEGKTSGEIACILGITERTVNYHL
NQVTRKTSNMRYQAIAGVSSGILLNLEQVVVTFNPKLMQ"

gene 2908..3486
/gene="yggA"
CDS 2908..3486
/gene="yggA"
/note="similar to Escherichia coli YggA"
/codon_start=1
/transl_table=11
/product="yggA"
/protein_id="AAB70019.1"
/db_xref="GI:1519238"

/translation="MIPIGAQNAFVLSRGHNRHLLAATLCCLDLILIGVFGG
ANLAAAPGIGALTWGVLPFCWFGIRSLRSAMQGOQAAALADSPRLMGVKSVMATL
GVTLINPHVYLDITLMLGSGFQSPAEPLRPAFAAGAMLASLVNPFYSLAFGAALSPWL
ARGVQQAIDTIVGLINGLALQALASGALLAS"

gene 3614..4033
/gene="cdpB"
CDS 3614..4033
/gene="cdpB"
/note="CdpB; similar to Escherichia coli CdpB"
/codon_start=1
/transl_table=11
/product="2'-3'-cyclic phosphodiesterase"

/protein_id="AAB70020.1"
/db_xref="GI:1519239"
/translation="MKLGATAIIVLLSACSNENDSKTPGAVALLRIQTSDIHSNVLG
YDYONKEDRKFGLSRTALLIRAAARSENPNLLDNDGLLQGTPLADYIFEQSGAGYL
DKOAHVPFKANVELGYDAGNLGNHEFNGLDYLTKTLQ"

BASE COUNT 787 a 1160 c 1229 g 857 t
ORIGIN

Query Match	8.7%	Score 61.6;	DB 1;	Length 4033;
Best Local Similarity	52.3%	Pred. No. 0.0001;		
Matches 136;	Conservative 0;	Mismatches 124;	Indels 0;	Gaps 0;
Qy 437	TGTTGATGGCAATCGTGTGACCTGGTTGAACCCGAATGCGTATTGACCGGTTTGTTGT	496		
Db 3200	TGGCCATGACGCTCGGGGTGACCTGCTCAATCCTCACTCTATCTGATACCCCTGATGC	3259		
Qy 497	TTATCGGGGCTCGCGCGCAATACGCGCACCGGACGGTGGATTTCGCGCGTGGCG	556		
Db 3260	TGCTCGGCTCCTTTGGCAGCCAGTTTGGCGAAGCCTTGGCGCCGCCCTTTGGCGCGGGG	3319		
Qy 557	CGTTCGCGGCAAGCCTGATCTGGTTCCGCGCTGGTGGGTTTCGGCGCAGCAGCATGTGAC	616		
Db 3320	CCATGCTGGCTTCCCTGCTGTCTGTTCTACAGTCTCGCCTTTGGTGGCGCGGCTCTCAC	3379		
Qy 617	GCCCGCTGCCAGCCCGCAGGTGTGGCGTGGATCAAGTCGTCGTGGAGTTGTGATGA	676		
Db 3380	CTGCTGCTTGCACGGGCGGGGTGCAGAGCTATTGATCTATTGTTGTTAATCATGT	3439		
Qy 677	CGCATTTGCCATCAAACTG	696		
Db 3440	TGGGACTGSCATTGCAACTG	3459		

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
C 1	60.4	8.5	830	17	CNS01MGH	AL151258	Anopheles
2	48.6	6.8	925	17	CNS0091P	AL053013	Drosophila
C 3	44.4	6.2	644	14	BQ704105	BQ704105	946110C04
C 4	42.8	6.0	532	10	AW927407	AW927407	945001E08
5	41.2	5.8	910	17	CNS0060N	AL065629	Drosophila
6	40.2	5.7	613	10	BE498441	BE498441	WHE0967 G

FEATURES
source Location/Qualifiers
1..830 /organism="Anopheles gambiae"
/strain="PEST"
/db xref="taxon:7165"
:/clone="22E24"
/clone_lib="NotreDame1"
/notes="end : T7"
BASE COUNT 207 a 231 c 227 g 156 t 9 others
ORIGIN

Query Match 8.5%; Score 60.4; DB 17; Length 830;
Best Local Similarity 50.78; Pred. No. 5.4e-07;
Matches 142; Conservative 1; Mismatches 137; Indels 0; Gaps 0;

QY 23 TTACAGGCTCTCTTTTGGGGCCAGTCCTTTTACTGTCTCATCGGACCAGCAATGTACTGG 82
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 KTC AAGGCTTGCCCTTAGTGCATTTTTATTCTTCCCCTCGGCCGCAAATGCGTTTCG 538

QY 83 TGATTAAACAAGGAATTAAGCGGAAGACATCATTCGGTCTTCTCTCGTGTCTTAATTT 142
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 537 TGAATCAACAGGCATCCGCCGCCAGTACCATCTGATGATGCCCTGCTATGCGCGGTAA 478

QY 143 CTGACGCTCTTTTGTTCATCGCCGCCACCTTGGCGGTTGATCTTTTGTCCAATGCCGCGC 202
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 GTGATTTGCTGTCATCTGTGCGGGATTTTGGCGGCGAGCGCTGCTGATGCAGTCTC 418

QY 203 CGATCGTCTCGAATATATGCTGGGTGGATCGCTTACTCTGTATGTTTGGCGGTCA 262
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 417 CGTGCTCTCTGCGGTGTCTACCTTGGCGGCGGTGCGCTTCTGCTGTACGAGATTTCG 358

QY 263 TGGCAGCGAAAAGACGCCCATGACAAACAAGGTGGAAGCGCC 302
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 GCGCGCTGAAAACGGCGATGAGCAGTAACTCGAACTGCG 318

RESULT 2
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TENT3 end of BAC #
DEFINITION BACR19D16 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL053013
AL053013.1 GI:4934461
GSS.

SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.

REFERENCE Direct Submission
AUTHORS Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ;
TITLE BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)
JOURNAL - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
Location/Qualifiers
1...925

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

```

FEATURES
  source
    Location/Qualifiers
      1..910
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BACR14J21"
        /clone_lib="RPCI-98"
        /note="end : 17"
BASE COUNT      202 a      63 c      112 g      198 t      335 others
ORIGIN

Query Match      5.8%; Score 41.2; DB 17; Length 910;
Best Local Similarity 14.8%; Pred. No. 0.35;
Matches 47; Conservative 150; Mismatches 119; Indels 2; Gaps 1;

```

Qy 333 GCCCGATGACACGGCTTTTGGCGGTTCGGCGGTGCACACTGACACGCCGAACCCGGGTGCG 392
| : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : :
Db 518 GSSSKKGGCGSGYGGGGYGVTGSGSGGGSSSSSCGSSSSSSSSSSSSSSSSSSSSSSSGBGR 577

[illegible]

393 GGTGGAGTGAGCGTCCATAAGCAGCGGGTTTGGSTAAAGCCCATGTTGATGGCAATCGT 452

Db 578 SSCGCGSCYSSSSTGTCGCGSSSSSSSSSSSSSSSTSTSSSSSTSGCGYSSSGSBSSC 637

[illegible]

Db 638 SSTSTSSSTTBETBKTSTYSTSTSTSTSTSTSTSTSTSTSTSTSTSTSS 697

Qy 511 GGCGGCAATACGGCGACACCGGACGGTGGATTTCGCCGCTGGCGCGTTCGGGGCAAGC 570

```

Db      698 SSSSSSTCKBYSTBSYSBITBTBTSTSTBBTSTBYTBBTBTYKSTTSMTST 757
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : :

```

571 CTGATCTCGGTTCCCGCTGGTGGGTTTCCGCGCAGCAGCATTTCTGCA CCCCCGGTGTCCAGC 630

d6	750	VTTTGGGCTGTGTCATCGAGCATTGTCACGCCCGCCGCTGTCCAGC	833
QY	571	CIGAHCTGGTTCCTCCGCCTGGTGTTCTGGCGCAACAGCATGTAC	

758 YTTBBSTKBSHTBSFKSTGKTGKTBRSBTTCTSSSSSBSTSYSTSCBSBSBSSTSY 817

QY 631 CCAAGGTGTGGCGCTGG 648

Db 818 SBCTSTSTSSBBSBS 835

RESULT 6	BE498441	BE498441	613 bp	mRNA	linear	EST 04-AUG-2000
LOCUS	WHE0967_G03	M05ZS	Wheat	pre-anthesis	spike cDNA library	Triticum
DEFINITION	aestivum	cDNA clone	WHE0967	G03	M05,	mRNA sequence.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BE498441	BE498441.1	GI:9697058	bread wheat.	Triticum aestivum
BE498441	BE498441.1	GI:9697058	bread wheat.	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
BE498441	BE498441.1	GI:9697058	bread wheat.	: Triticeae; Triticum

REFERENCE
AUTHORS
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J., Seaton, C.L. and Tong, J.C.
TITLE
The structure and function of the expressed portion of the wheat genomes - Pre-anthesis spike cDNA library
JOURNAL
Unpublished (2000)

WORKSHEET
 COMMENT
 Unpublished (2000)
 Contact: Olin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific
 West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: caandersn@pw.usda.gov
 Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 See www.Steelhead.org subject

```

FEATURES ,
source
seq primer: StrataGene SK primer.
Location/Qualifiers
1..613
/organism="Triticum aestivum"

```

```

/cultivar="Chinese Spring"
/db_xref="taxon:4585"
/clone="WHE0967_G03_M05"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLA"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pBluescript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
144 a 178 c 188 g 102 t 1 others

```

	Query Match	5.7%	Score 40.2;	DB 10;	Length 613;
	Best Local Similarity	53.5%;	Pred. No. 0.61;		
	Matches 84;	Conservative 0;	Mismatches 73;	Indels 0;	Gaps 0
Qy	380	GCAACCGGTCGGGTGAGGTGACGTCGATAGCAGCGGTTTGGTAAAGCCCATGT	439		
Db	84	GCBAAGGATGCGGCGCAGAGGGGAGGCGCATCCGCGTGTCCGCGTCCGGCGGACGG	143		
Qy	440	TGATGCCAATCGTGTGACCTGGTTTGAAACCCGAATGGGTATTTTGACGCGTGTGTGTTTA	499		
Db	144	TGATGCAGATCTGCGGGACCGCGGTAACCTGGTAGTGACCGCGACGTCGACATGACCA	203		
Qy	500	TCGGCGGCGTCGGCGGCGCAATACGGCGACACCGAAG	536		
Db	204	TGGCGGAGTTCGTGGGCGGTACGGCGACCCCGTGCG	240		

RESULT 7
CNS0091P/c

CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BAC19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)

GENOSCOPE.	Genoscope - Centre National de Sequencage :
Direct Submission	Submitted (02-JUN-1999) Genoscope - FRANCE (E-mail : segref@genoscope.cns.fr)
TITLE	BP 191 g1006c EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)
JOURNAL	- Web : www.genoscope.cns.fr
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamooser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
FEATURES	Location/Qualifiers

```
source
1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPI-98"
/note="end : TET3"

BASE COUNT      120 a      61 c      61 g      172 t      511 others
ORIGIN

Query Match      5.6%; Score 40; DB 17; Length 925;
Best Local Similarity 13.7%; Pred. No. 0.81;
Matches 51; Conservative 157; Mismatches 164; Indels 0; Gaps 0;

QY 197 CCGCCCGCATCGTCTGATATATGCTGGGTGGCATGCTTACCTGTTATGCTTTG 256
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 925 CSBSCSCSCSCSBSCSSSMTSSNSBSCSSBSSSSTSSMSBSSSSSSSSSS 866

QY 257 CCCTCATGGCAGCAAGACCCATGACAAACAGGTGGAGCCGACAGATATTGAAG 316
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 865 SSGTSSACVKCNASSSCGCCGCMABCWCSSSSSCGGSASARGVKVRASGAGKRGGS 806

QY 317 AAACAGAACCAACCGTCCCATGACACGCTTTGGGGGCTTCGGCGTGCCTGACA 376
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 805 GGASASHSSSACBSSSSSCASCMWSASSSSASRSRSGGAGGSCASSRSSSSSS 746

QY 377 CGCGCAACCGGGTGGGTGAGTGACGCTCGATAAGCAGCGGTTGGGTAAAGCCCA 436
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 745 ASAGSVVSSASSSSSSSSSVSCSVASSMSBSSSSASASSSSSSSSSASCSCCC 686

QY 437 TGTTCATGGCAATCGTCTGATGCTGTGTAACCGAATGCTATTGACGCGTTGTGT 496
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 685 TSWSCSTASMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 626

QY 497 TTATCGCGCGCTCGCGCGCAATACGCGACACCGAGCGTGGATTTCGCGCTGGCG 556
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 625 SMSSGSGSVSSAGSMSSSVSSGSGSGGGGCGGSGSSSSSGSGSGSGSVCS 566

QY 557 CGTTCCGCGCAA 568
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 565 SGCRCSCSSA 554

RESULT 8
AY108197
LOCUS      AY108197
DEFINITION Zea mays PC0065133 mRNA sequence.
ACCESSION  AY108197
VERSION     AY108197.1 GI:21211275
KEYWORDS   HTC.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 1667)
AUTHORS     Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S.,
            Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
JOURNAL     Submitted (25-APR-2002) Maize Mapping Project, University of
REFERENCE   2 (bases 1 to 1667)
AUTHORS     Coe, E.C.
            Direct Submission
            Submitted (25-APR-2002) Maize Mapping Project, University of
JOURNAL     Missouri, Columbia, MO 65211, USA
FEATURES    Location/Qualifiers
            1. .1667
            /organism="Zea mays"
            /db_xref="MaizeDB:633502"
            /db_xref="taxon:4577"
            /clone="PC0065133"
            /clone_lib="Maize Mapping Project/DuPont Consensus
            Library"

source
1. .1667
/organism="Zea mays"
/db_xref="MaizeDB:633502"
/db_xref="taxon:4577"
/clone="PC0065133"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"

BASE COUNT      215 a      303 c      246 g      186 t
ORIGIN

/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT      361 a      490 c      482 g      334 t
ORIGIN

Query Match      5.6%; Score 39.6; DB 11; Length 1667;
Best Local Similarity 47.2%; Pred. No. 1.3;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 405 CGTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTT 464
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 979 CGCGCGCGCGCAGCTGACTACGTGCGCGGCGACCATCAGCGCGCTGGAGAGTCTGTGA 1038

QY 465 GAACCCGAATCGTATTGAGACGCGTTTGTATTATCGCGCGCTCGCGCGCAATACGG 524
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1039 GCGCGCGCGCGCGCAGCTCCCGTGTTCGTGACGCGCGCGTCCGCGCGGACCGT 1098

QY 525 CGACACCGGACGCTGATTTTCCTCCCTCGCGGTTTCGCGGCAAGCTGATCTGTTCC 584
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1099 CTTCAAGCGCTCGCGCTCGCGCGCGCGTATTTCGTGCGGAGCGCGTGTCTC 1158

QY 585 GCTGTGTGGTTTCGCGCGCAGCAGCATTTGTCACGCGCTGTCAGCCCAAGGTGTCGG 644
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1159 CTGCGCGCGCGCGCGCGCGCGCTGTCACGCTCTCGCGCATCTCGCGGACGAGTT 1218

QY 645 CTGGATCAACGCTG 658
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1219 CGAGCTCACCATGG 1232

RESULT 9
AY104298/c
LOCUS      AY104298/c
DEFINITION Zea mays PC0095355 mRNA sequence.
ACCESSION  AY104298
VERSION     AY104298.1 GI:21207376
KEYWORDS   HTC.
SOURCE      Zea mays.
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 950)
AUTHORS     Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S.,
            Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
JOURNAL     Submitted (25-APR-2002) Maize Mapping Project, University of
REFERENCE   2 (bases 1 to 950)
AUTHORS     Coe, E.C.
            Direct Submission
            Submitted (25-APR-2002) Maize Mapping Project, University of
JOURNAL     Missouri, Columbia, MO 65211, USA
FEATURES    Location/Qualifiers
            1. .950
            /organism="Zea mays"
            /db_xref="MaizeDB:635429"
            /db_xref="taxon:4577"
            /clone="PC0095355"
            /clone_lib="Maize Mapping Project/DuPont Consensus
            Library"

/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT      215 a      303 c      246 g      186 t
ORIGIN
```

```

Query Match      5.5%; Score 39; DB 11; Length 950;
Best Local Similarity 46.5%; Pred. No. 1.6;
Matches 166; Conservative 0; Mismatches 185; Indels 6; Gaps 1;

QY 349 TTGGCGGTTTCGGCGTGGCCACTGACACCGCAACCGGTGGGGTGGAGGTGAGCGTC 408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 TTGGCCCTGGACCGCGGAGAGTCCGATGGCGCCCTGGAGAGCTTGCCTGGTTATGGTC 333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 409 GATAAGCAGCGGGTTCGGGTAAGCCCATGTTGATGCAATCGTCTGACCTGGTTGAAC 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 CTCTCGTAGTAGCAGAGTGGGAGCCGAGTGGAGCTGCTGGCGAGTGGATC 273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 CCGAATCGGTATTGGACGGCTTTGTGTTTATCGCGCGGTGCGCGCGCAATACGCGCAC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 TCGAAGCGCCCGTGGCTCGTTCGAAGCTGTAGGAGTGACGAGTCCGGGATGAGCCCC 213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 ACCGACGCGTGGATTTTCGCGCGTGGCGCTTCGCGCAAGCCTGATCTGTTCCCGCTG 588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 TTGGGAGCGGCTACTTGGGACAGAGTCTGTTCCCGCGCGCTGAGGAGCGC-----G 159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 GTGGGTTTCGCGCAGCAGCATGTTCAAGCCGCTGTCCAGCCCGCCAAAGGTGTGCGCTGG 648
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 GCGACGCGTGGCGCGCACGCGCGCGGAGGAGGACGAGGACACGAGCGCGCGCGG 99
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 ATCAACGCTGCTGGCGAGTGTGATACCGCATGTCGCGCATCAAACTGATGTTGATG 705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 GGCAGCGGAGACATGGCGCTTTCGCACTTTCGAGTTTCGGGGAAGAGTTCCTGGTG 42
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AW787854/c
LOCUS      AW787854
DEFINITION 945004G09.X1 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION  AW787854
VERSION     AW787854
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 557)

REFERENCE   1 (bases 1 to 557)
AUTHORS    Walbot, V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL    Unpublished (1999)
COMMENT    Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945004 row: G column: 09.
FEATURES    Location/Qualifiers
             1..557
              /organism="Zea mays"
              /cultivar="W23"
              /db_xref="taxon:4577"
              /clone_lib="945 - Mixed adult tissues from Walbot lab,
              same as 707 (SK)"
              /tissue_type="tassel, kernal, silk, husk, root, leaf"
              /dev_stage="fully-grown"
              /lab_host="DH10B"
              /note="Organ: tassel, kernal, silk, husk, root, leaf;
              vector: pBAD10; Site_1: EcoRI; cDNA library from fully
              differentiated maize tissues from an active Mutator plant.
              Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,
              root, leaf). Unidirectionally cloned. New library number
              given to library 707 for additional sequencing."

BASE COUNT      81 a 215 c 161 g 100 t

```

ORIGIN

```

Query Match      5.5%; Score 38.8; DB 10; Length 557;
Best Local Similarity 48.2%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 349 TTGGCGGTTTCGGCGTGGCCACTGACACCGCAACCGGTGGGGTGGAGGTGAGCGTC 408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 TTGGCCCTGGACCGCGGAGAGTCCGAGATGGCGCCCTGGAGAGCTTCCCGGTTATGGTC 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 409 GATAAGCAGCGGGTTCGGGTAAGCCCATGTTGATGCAATCGTCTGACCTGGTTGAAC 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 CTCTCGTAGTAGCAGAGTGGGAGCCGAGTGGACGTCGACAGTGTGCGGAGGTGGATC 212
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 CCGAATCGGTATTGGACCGCTTTGTGTTTATCGCGCGCTGCGCGCGCAATACGCGCAC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 TCGAAGCGCCCGTGGCTCGTTCGAAGCTGTAGGAGTGCAGGAGTCCGGGATGAGCCCC 152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 ACCGACGCGTGGATTTTCGCGCGTGGCGCTTCGCGCAAGCCTGA 574
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 TTGGGAGCGCTACTTGGGACAGAGTTCGTTCCGCGCGCGCTGA 106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 11

```

BG975548/c
LOCUS      BG975548
DEFINITION 602845228F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:498095 5',
            mRNA sequence.
ACCESSION  BG975548
VERSION     BG975548.1 GI:14363185
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
            1 (bases 1 to 903)

REFERENCE   1 (bases 1 to 903)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10981 row: g column: 08
            High quality sequence stop: 250.
FEATURES    Location/Qualifiers
             1..903
              /organism="Mus musculus"
              /strain="NMRI"
              /db_xref="taxon:10090"
              /clone="IMAGE:498095"
              /clone_lib="NCI CGAP Mam4"
              /tissue_type="tumor, gross tissue"
              /dev_stage="5 months"
              /lab_host="DH10B"
              /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
              Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
              Library constructed by Life Technologies. Investigators
              providing samples: Lothar Hennighausen/Priscilla Furth,
              NIH Reference for transgenic model: Li et al., Cell Growth
              and Differentiation 7, 3-11 (1996)."

BASE COUNT      373 a 167 c 297 g 66 t

```

FEATURES

```

source
Location/Qualifiers
1..903
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:498095"
/clone_lib="NCI CGAP Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7, 3-11 (1996)."

BASE COUNT      373 a 167 c 297 g 66 t
ORIGIN
Query Match      5.3%; Score 38; DB 13; Length 903;
Best Local Similarity 55.2%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

```

QY 127 CTCGTGCTTTAATTTCTGACGCTCTTTTGTTCATCGCGGCACCTTGGCGGTGATCTT 186
Db 560 CTCGTGCTCTAGTCGCGTCTTGTCGCGTCCGTCGCTTGTGTCGTCGCGTC 501
QY 187 TTGTCATGCGCGCGGATCGTCTGATATATGATGCTGCGGTGGCATCGCTTACCTG 246
Db 500 TGCTGCTCGCTGCGCGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 441
QY 247 TTATGTTTGGCGT 260
Db 440 TTTCTGTTTGTGCT 427

RESULT 12
LOCUS CENS01H1J 804 bp DNA linear GSS 01-JUN-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 08N07 of Notredame library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL144476
VERSION AL144476.1 GI:7002638
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 804)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BMJ, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.

FEATURES
source
1..804
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="08N07"
/notes="end : SP6"
BASE COUNT 181 a 257 c 207 g 148 t 11 others
ORIGIN
Query Match 5.3%; Score 37.8; DB 17; Length 804;
Best Local Similarity 48.1%; Pred. No. 3.6;
Matches 137; Conservative 1; Mismatches 143; Indels 4; Gaps 1;

QY 412 AAGCAGCGGTTGGTAAAGCCCATGTTGATGCGATCGTGTGACCTGTTGACCCG 471
Db 660 ATGAAGCAGGCGCGTGGAAAGATTATCGTCACCATGCTGCGGTGACCTGGCTTAACCCG 601
QY 472 AATGCGTATTGACGCGTTGTGTTTATCGCGGCTGCGCGCAATACGCGACACC 531
Db 600 CATGTGATCTCGACACTTTCGTGCTGGCAGCTGGCGCGGTCACTGGAGGTTGAG 541
QY 532 GACGGTGAATTTTCGCGCTGCGGCTTCGCGCAAGCTGATCTGTTTCCGCTGGTG 591
Db 540 CCGAAACGCTGGTTTGGCTCGGTACGCTCAGCGCTCTTCTCTGTTGTTCTTCGCTC 481
QY 592 GGTTCGCGGCAGCATGTCACGCGGCTGTCAGCGCCCTGTCAGCGCCCAAGGTGGCGCTGCATC 651
Db 480 CATTCGCGGCGCATG----TGGCGCGCGCGCTGCTGTCAGCGCCCAAGCGCGCATCAT 425

```

```

QY 652 AACGTGCTGTCAGTGTGATGACCGCATTTGGCCATCAAACTG 696
Db 424 AATACCTGTCGCGCTGGTATGTTATCGCTTTCAGCTG 380

RESULT 13
LOCUS BJ295166/c 370 bp mRNA linear EST 09-APR-2002
DEFINITION BJ295166 Y. Ogihara unpublished cDNA library, Wh_SL Triticum
aestivum cDNA clone whs12a03 3', mRNA sequence.
ACCESSION BJ295166
VERSION BJ295166.1 GI:20109847
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 370)
AUTHORS Ogihara,Y. and Murai,K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1..370
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whs12a03"
/tissue_type="seed DPA30"
/dev_stage="Peekes' scale 11.3"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhunov
in J Dvorak Lab). Total RNA was prepared from sheath
tissue, equal quantities of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo
excised to give pBluescript phagemids in the TJ Clouse lab
at the University of California, Riverside (Akhunov, Chin
, Choi, Clouse, Fenton, Kianian, Otto, Simons, Zhang).
Plasmid DNA preparations and DNA sequencing were
performed in the Od Anderson lab (all other authors)."
BASE COUNT 63 a 127 c 85 g 93 t
ORIGIN
Query Match 5.3%; Score 37.6; DB 13; Length 370;
Best Local Similarity 50.0%; Pred. No. 3.1;
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 264 GGCAGCGAAAGACGCATGACAAACAGGTGAAGCGCCACAGATCATTTGAAGAACAGA 323
Db 368 GCGCGCCAGAGACGCATGTTGGAGAGACGCGCTCGGCCAGAGACGCCGCGATGGAGAC 309
QY 324 ACCAACCGTGCCTGATGACGCTTTTCGGCGGTTTCGGCGGTGCGCACTGACACGCGCAA 383
Db 308 GCGCGAGCGCGCAAGGACACGCGCTGGGAGACGCGCGGCGGCGCAAGGATAAGGCCAA 249
QY 384 CCGGTGCGGTGGAGGTGAGCGTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGAT 443
Db 248 CCAGGGGTACGAGAAAGTGAAGGAGAGGTTTGGGAGGTGAAGGACAAAGGTCAACCGGCG 189
QY 444 GGCATCG 451

```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 17:06:23 ; Search time 156.463 Seconds

(without alignments)

10233.535 Million cell updates/sec

Title: US-09-105-117k-1_COPY_1016_1726

Perfect score: 711

Sequence: 1 atgggtgatcatgaaatctt...-.....aactgatgttgatgggttag 711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 101002:*

1: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	711	100.0	822	22	AAAF71779
2	711	100.0	822	23	AAAF71779
3	711	100.0	2374	18	AAAT96816
4	711	100.0	349980	22	AAHG8528
5	708	99.6	708	22	AAHG8420
6	317.6	44.7	1568	22	AAH45375
7	60.4	8.5	636	21	AAAS2691
8	57	8.0	15239	17	AAAT3536
9	57	8.0	4403765	22	AAI99683

10	57	8.0	4411529	22	AAI99682
11	47.4	6.7	7521	23	AAAT9682
12	47.4	6.7	7521	23	AAAT9682
13	39	5.5	993	22	AAAF71777
14	39	5.5	993	23	AAAF71777
15	37.6	5.3	4403765	22	AAI99683
16	37.6	5.3	4411529	22	AAI99682
17	37	5.2	1155	22	AAAT9682
18	36.8	5.2	1920	23	AAAT9682
19	36.8	5.2	6029	23	AAAT9682
20	36.8	5.2	6454	23	AAAT9682
21	36.8	5.2	14395	23	AAAT9682
22	36.6	5.1	2842	21	AAAT9682
23	36	5.1	1068	22	AAAT9682
24	36	5.1	349980	22	AAAT9682
25	35	4.9	5828	21	AAAT9682
26	34.8	4.9	7158	24	AAAT9682
27	34.6	4.9	423	23	AAAT9682
28	34.6	4.9	6611	17	AAAT9682
29	34.6	4.9	6611	18	AAAT9682
30	34.6	4.9	6611	18	AAAT9682
31	34.6	4.9	6611	18	AAAT9682
32	34.4	4.8	459	22	AAAT9682
33	34.4	4.8	3318	15	AAAT9682
34	34.4	4.8	4372	18	AAAT9682
35	34.2	4.8	125401	22	AAAT9682
36	34	4.8	300	20	AAAT9682
37	34	4.8	515	20	AAAT9682
38	34	4.8	2048	21	AAAT9682
39	34	4.8	5224	20	AAAT9682
40	34	4.8	524	22	AAAT9682
41	34	4.8	5488	22	AAAT9682
42	33.8	4.8	2005	24	AAAT9682
43	33.8	4.8	2005	24	AAAT9682
44	33.8	4.8	2024	21	AAAT9682
45	33.8	4.8	2181	22	AAAT9682

ALIGNMENTS

RESULT 1

AAAF71779

ID AAAT9682 standard; DNA; 822 BP.

XX AAAT9682

AC AAAT9682

XX 30-APR-2001 (first entry)

DT AAAT9682

XX AAAT9682

DE AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

XX AAAT9682

PD 13-SEP-2001.
 PF 22-DEC-2000; 2000WO-IB02035.
 PR 09-MAR-2000; 2000US-187970P.
 PR 23-JUN-2000; 2000US-0606740.
 XX
 PA (BADI) BASF AG.
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
 PI Lee H, Hwang B;
 XX
 DR WPI; 2001-582269/65.
 DR P-PSDB; AAU71888.
 XX
 PT Nucleic acids encoding metabolic pathway proteins from *Corynebacterium*
 PT glutamicum, useful for producing methionine and lysine in
 PT *Corynebacterium* and *Brevibacterium* -
 XX
 PS Disclosure; Page 214-215; 316pp; English.
 XX
 CC The present invention relates to the isolation of novel *Corynebacterium*
 CC glutamicum genes encoding metabolic pathway (MP) proteins
 CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention
 CC include enzymes involved in the lysine and methionine biosynthetic
 CC pathways. The polynucleotide sequences of the invention can be used
 CC for the large-scale production and/or modulation of expression of
 CC fine chemicals such as lysine and methionine. The sequences of the
 CC invention may be used to identify *C. glutamicum* and related organisms
 CC e.g. *C. diphtheriae* in a subject to detect diphtheria.
 CC AA996073-AAA996132 represent *C. glutamicum* genes encoding the novel
 CC metabolic pathway proteins of the invention.
 XX
 SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 100.0%; Score 711; DB 23; Length 822;
 Best Local Similarity 100.0%; Pred. No. 2.8e-206;
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 60
 DB 92 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 151
 QY 61 ATCGGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGGAAGGACTCAATCGG 120
 DB 152 ATCGGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGGAAGGACTCAATCGG 211
 QY 121 GTTCTTCTGCTGTTAAATTTCTGACGCTTTTGTTCATGCGCCGCACTTGGCGGTT 180
 DB 212 GTTCTTCTGCTGTTAAATTTCTGACGCTTTTGTTCATGCGCCGCACTTGGCGGTT 271
 QY 181 GATCTTTTGTCCAAATGCCGCGCATGCTCGATATTATGCGTGGGTGGCATCGCT 240
 DB 272 GATCTTTTGTCCAAATGCCGCGCATGCTCGATATTATGCGTGGGTGGCATCGCT 331
 QY 241 TACCTGTTATGTTTCCGCTCATGGCAGAAAGCGCATGACAAACAAAGTGAAGCG 300
 DB 332 TACCTGTTATGTTTCCGCTCATGGCAGAAAGCGCATGACAAACAAAGTGAAGCG 391
 QY 301 CCACAGATCATTTGAAGAAACAGAAACAAACCGTCCCGATGACACGCTTTGGCGGTTTCG 360
 DB 392 CCACAGATCATTTGAAGAAACAGAAACAAACCGTCCCGATGACACGCTTTGGCGGTTTCG 451
 QY 361 GCGGTGGCCACTGACACCGGCAACCGGTGCGGTGGAGGTGAGCGTGCATAGCAGCGG 420
 DB 452 GCGGTGGCCACTGACACCGGCAACCGGTGCGGTGGAGGTGAGCGTGCATAGCAGCGG 511
 QY 421 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGCTGACCTGTTCAACCCGAATGCGTAT 480
 DB 512 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGCTGACCTGTTCAACCCGAATGCGTAT 571
 QY 481 TTGAGCGGTTGTGTTTATCGCGCGCTCGCGCGCAATACGGCGACACCGGACGGTGG 540

DB 572 TTGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACGGCGACACCGGACGGTGG 631
 QY 541 ATTTTCGCGCTGGCGGTTTCGCGCAAGCTGATCTGTTCCGCTGGTGGTTCGGC 600
 DB 632 ATTTTCGCGCTGGCGGTTTCGCGCAAGCTGATCTGTTCCGCTGGTGGTTCGGC 691
 QY 601 GCAGCAGCATTTGTCAACCGCGCTGTCCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660
 DB 692 GCAGCAGCATTTGTCAACCGCGCTGTCCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 751
 QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCCATCAAACTGATGTGATGGGTTAG 711
 DB 752 GTGGCAGTTGTGATGACCGCATTTGGCCCATCAAACTGATGTGATGGGTTAG 802

RESULT 3
 AAT96816
 ID AAT96816 standard; DNA; 2374 BP.
 XX
 AC AAT96816;
 XX
 DT 12-MAR-1998 (first entry)
 XX
 DE DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.
 XX
 KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
 KW Microbial production; amino acid; animal feed additive; ds.
 XX
 OS *Corynebacterium glutamicum*.
 XX
 FH Key Location/Qualifiers
 FT CDS complement (82..954)
 FT /*tag= a
 FT /label= LysG
 FT 1016..1726
 FT /*tag= b
 FT /label= LysE
 FT complement (1723..2373)
 FT /*tag= c
 FT /label= orf3
 XX
 PN DE19548222-A1.
 XX
 PD 26-JUN-1997.
 XX
 PF 22-DEC-1995; 95DE-1048222.
 XX
 PR 22-DEC-1995; 95DE-1048222.
 XX
 PA (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 PI Eggeling L, Sahn H, Vrljic M;
 XX
 DR WPI; 1997-333867/31.
 DR P-PSDB; AAM37714-16.
 XX
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT *Corynebacterium*
 XX
 PF Claim 23 and 26; Page -; 16pp; German.
 XX
 CC This DNA, isolated from *Corynebacterium glutamicum*, contains the LysG,
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine,
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.

DB 128533 GTTGGGTAAGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGAATCGGTAT 128474
 QY 481 TTGGAGCGCTTTGTGTTTATCGCGCGCTCGCGCGCATACGGCGACACCGAGCGTGG 540
 DB 128473 TTGGAGCGCTTTGTGTTTATCGCGCGCTCGCGCGCATACGGCGACACCGAGCGTGG 128414
 QY 541 ATTTTTCGCGCTGGCGGCTTCGCGCAAGCTGATCTGGTTCCTCCGCTGGTGGGTTTCGCG 600
 DB 128413 ATTTTTCGCGCTGGCGGCTTCGCGCAAGCTGATCTGGTTCCTCCGCTGGTGGGTTTCGCG 128354
 QY 601 GCAGCAGCATTTGTCAGCCCGCTGTCAGCCCAAGGTGTGGCGTGGATCAACGTCGTC 660
 DB 128353 GCAGCAGCATTTGTCAGCCCGCTGTCAGCCCAAGGTGTGGCGTGGATCAACGTCGTC 128294
 QY 661 GTGGCAGTTGTGATGACCCCATTTGGCCATCAAACTGATGTTGATGGGTTAG 711
 DB 128293 GTGGCAGTTGTGATGACCCCATTTGGCCATCAAACTGATGTTGATGGGTTAG 128243

RESULT 5

AAH68420

ID AAH68420 standard; DNA; 708 BP.

XX AC AAH68420;

XX DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 3455.

XX DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX FN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX DR P-PSDB; AAG93201.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying

XX FT mutation point of a gene, measuring expression of a gene, analysing

XX FT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 1; SEQ ID NO: 3455; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein

XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

XX CC are useful for identifying the mutation point of a gene derived from a

XX CC mutant of coryneform bacterium, measuring expression amount and

XX CC analysing the expression profile or expression pattern of a gene derived

XX CC from Coryneform bacterium, and identifying a homologue of a gene derived

XX CC from Coryneform bacterium. Coryneform bacteria are useful for producing

XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

XX CC particularly L-lysine. The present sequence is a nucleic acid described

XX CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from the

XX CC European Patent Office.

XX SQ Sequence 708 BP; 134 A; 173 C; 221 G; 180 T; 0 other;
 Query Match 99.8%; Score 708; DB 22; Length 708;
 Best Local Similarity 100.0%; Pred. No. 2.1e-205;
 Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGCTGATCATGGAATCTTTTATTACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 60
 DB 1 ATGCTGATCATGGAATCTTTTATTACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 60
 QY 61 ATCGGACCGCAGAAATGACTGTGATTAACAAGAGAAATTAAGCGCGAGGACTCATTTGCG 120
 DB 61 ATCGGACCGCAGAAATGACTGTGATTAACAAGAGAAATTAAGCGCGAGGACTCATTTGCG 120
 QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGTCTTTTGTTCATCCCGGACCTTTGGGCGTT 180
 DB 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGTCTTTTGTTCATCCCGGACCTTTGGGCGTT 180
 QY 181 GATCTTTTGTCCAAATGCGCGCGATCGTCTGATATTATGCTGGGTGGCATCGCT 240
 DB 181 GATCTTTTGTCCAAATGCGCGCGATCGTCTGATATTATGCTGGGTGGCATCGCT 240
 QY 241 TACCTGTTATGTTGCTTCATGGCAGCGAAAGAGCCCATGACAAAGGTGGAAGCG 300
 DB 241 TACCTGTTATGTTGCTTCATGGCAGCGAAAGAGCCCATGACAAAGGTGGAAGCG 300
 QY 301 CCACAGATCAATTGAAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTGCG 360
 DB 301 CCACAGATCAATTGAAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTGCG 360
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTTCGATAGCAGCGG 420
 DB 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTTCGATAGCAGCGG 420
 QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGAATGCGTAT 480
 DB 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGAATGCGTAT 480
 QY 481 TTGGACGCGTTTGTGTTATCGCGCGGCTGCGCGCGCAATACCGCGACACCGGCGTGG 540
 DB 481 TTGGACGCGTTTGTGTTATCGCGCGGCTGCGCGCGCAATACCGCGACACCGGCGTGG 540
 QY 541 ATTTTCGCGCTGCGCGGTTTCGCGCAACCTGATCTGTTCCCGTGGGTTCGGC 600
 DB 541 ATTTTCGCGCTGCGCGGTTTCGCGCAACCTGATCTGTTCCCGTGGGTTCGGC 600
 QY 601 GCAGCAGCATTTGTCAGCCCGCTGTCAGCCCGCTGTCAGCGCTGGGATCAACGTCGTC 660
 DB 601 GCAGCAGCATTTGTCAGCCCGCTGTCAGCCCGCTGTCAGCGCTGGGATCAACGTCGTC 660
 QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGT 708
 DB 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGT 708

RESULT 6

AAH45375

ID AAH45375 standard; DNA; 1568 BP.

XX AC AAH45375;

XX DT 11-SEP-2001 (first entry)

XX C. thermoamniogenes lysin biosynthetic enzyme lyse DNA.

XX DE Heat-resistant; lysin biosynthesis; enzyme; coryneform;

XX KW aspartate-semialdehyde dehydrogenase; lyseE; ds.

XX OS Corynebacterium thermoamniogenes.

XX PN JP2001120270-A.

XX XX

QY 199 GCGCGCATCGTGTGATATATATGCGTGGGGTGGCATCGCTTACCTGTTATGTTTGGC 258
 Db |||||
 193 TCGCGGTGTTGCTGGCGCTGGTCACTGGGCGGCTAGCCTTCTTGTGTGTTATGTT 252
 QY 259 GTCATGGCAGCGAAGACGCGCATGACAAACAGGTGGA 296
 Db |||||
 253 TTTGGCGCTTTTAAACAGCAATGAGCAGTAATATTGA 290

RESULT 8
 AAT33536/c
 ID AAT33536 standard; DNA; 15239 BP.
 XX
 AC AAT33536;
 XX
 DT 15-FEB-1998 (first entry)
 XX
 DE BCG deletion region 2 and flanking sequences.
 XX
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;
 KW mycobacteria; vaccine; infection; marker; ss.
 XX
 OS Mycobacterium bovis strain BCG.

FH Key Location/Qualifiers
 FT misc_feature 3382..14071
 FT /*tag= a
 FT /note= "BCG delta 1 deletion region"

PN WO9625519-Al.
 XX
 PD 22-AUG-1996.
 XX
 PF 15-FEB-1996; 96WO-US01938.
 XX
 PR 17-FEB-1995; 95US-0390878.
 XX
 PA (PATH-) PATHOGENESIS CORP.
 PI Mahairas GG, Stover CK;
 DR WPI; 1996-393419/39.
 XX
 PT Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 PT bacterial infection
 XX
 PS Example 1; Fig 2; 66pp; English.

XX This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta2. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT33535 and AAT33537) have also been detected.
 CC Identification involved screening a BCG cosmid library with a
 CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.

SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;
 Query Match 8.0%; Score 57; DB 17; Length 15239;
 Best Local Similarity 53.1%; Pred. No. 2.1e-06;
 Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATGCTGCTGACCTGGTTGAACCCGATGCTATTGTCGCGCT 490
 |||||

Db 4792 AGGTCTCTGTGACCTGTGCGGCAATTCAGTTTCCTCAACCCACACGCTACCTCGACACCG 4733
 QY 491 TTGTGTTTATCGGCGCGCTCGGCGGCAATACGGCGACACCGGCGGTGGATTTTCGCCG 550
 Db |||||
 4732 TGTGTTGCTAGGCGCGCTGGCCAAACGAGCACAGCA---COAGCGCTGGGTGTTTCGGCC 4676
 QY 551 CTGGCGCGTTCGCGCAAGCCTGATCTGTTCCCGCTCGTGGTTCGCGCGCAGCAGCAT 610
 Db |||||
 4675 TCGGCGCGTTCACAGCCAGTGGGTATGTTGCGCACCTCGGTTTCGAGCGCGCGGT 4616
 QY 611 TGTCAAGCGCGCTGTCCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTCGTCGAGTTG 670
 Db |||||
 4615 TCGCGCGGCTGTTTCAACCAACCGCGCTCGTGGAGAACTCTCGACGCGCTGATCGCGTCA 4556
 QY 671 TGATGACCGCATTTGGCCATCAACTGATGTTGA 703
 Db |||||
 4555 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 4523

RESULT 9
 AAI99683
 ID AAI99683 standard; DNA; 4403765 BP.
 XX
 AC AAI99683;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 DR WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 PT 1551 and H37Rv differ
 PS
 XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
 Query Match 8.0%; Score 57; DB 22; Length 4403765;
 Best Local Similarity 53.1%; Pred. No. 2.9e-05;

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAAACCGAATGCGTATTTGGACGGT 490
Best Local Similarity 53.1%; Score 57; DB 22; Length 4411529;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAAACCGAATGCGTATTTGGACGGT 490
Db 2230334 AGTCTCTGGTGAACCTGTGCGGCATTCACGTTCTCAACCCACACGCTTACCTCGACACCG 2230393

QY 491 TTGCTGTTTATCGCGCGCTGCGCGCAATACGGCGCACACCGGCGTGGATTTTCGCCG 550
Db 2230394 TCGTGTGCTAGGCGCGCTGCGCAACGAGCAGCGA---CCAGCGTGGCTGTTCGCC 2230450

QY 551 CTGCGCGCTTCGCGCAAGCCTGATCTGTTCCCGCTGGTTCGCGCGCAGCAGCAT 610
Db 2230451 TCGCGCGCTGCACAGCAGTGGTATGTTTCGCCACCTCGCGTTCGAGCGCGGT 2230510

QY 611 TGTCAACCGCTGTCAGCGCCCAAGTGTGCGCTGGATCAAGTCTGTCGCGAGTTG 670
Db 2230511 TCGCGCGCTGTTCAACCAACCCCGCTGCGTGGAGAAATCTCGACGCGCTGATCGCGTCA 2230570

QY 671 TGATGACCGCATTCGCCCATCAAACTGATGTTGA 703
Db 2230571 TGATGTTGCGCTGGGAATCTCGTGAACCGTGA 2230603

RESULT 11
ID AAS71378
XX AAS71378 standard; cDNA; 7521 BP.
AC AAS71378;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #7182.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG07191.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 1; SEQ ID No 7182; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAAACCGAATGCGTATTTGGACGGT 490
Db 2227633 AGTCTCTGGTGAACCTGTGCGGCATTCACGTTCTCAACCCACACGCTTACCTCGACACCG 2227692

QY 491 TTGCTGTTTATCGCGCGCTGCGCGCAATACGGCGCACACCGGCGTGGATTTTCGCCG 550
Db 2227693 TCGTGTGCTAGGCGCGCTGCGCAACGAGCAGCGA---CCAGCGTGGCTGTTCGCC 2227749

QY 551 CTGCGCGCTTCGCGCAAGCCTGATCTGTTCCCGCTGGTTCGCGCGCAGCAGCAT 610
Db 2227750 TCGCGCGCTGCACGCAAGTGGTATGTTTCGCCACCTCGGTTTCGAGCGCGCGGT 2227809

QY 611 TGTCAACCGCTGTCAGCGCCCAAGTGTGCGCTGGATCAACGTCGTCGCGCAGTTG 670
Db 2227810 TCGCGCGCTGTTCAACCAACCCCGCTGCGTGGAGAAATCTCGACGCGCTGATCGCGTCA 2227869

QY 671 TGATGACCGCATTCGCCCATCAAACTGATGTTGA 703
Db 2227870 TGATGTTGCGCTGGGAATCTCGTGAACCGTGA 2227902

RESULT 10
ID AAI99682
XX AAI99682 standard; DNA; 4411529 BP.
AC AAI99682;
XX
DT 15-JAN-2002 (first entry)
XX
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN US6294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-0103840.
XX
PR 24-JUN-1998; 98US-0103840.
XX
PA (GENO-) INST GENOMIC RES.
XX
XX Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen.
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 7521 BP; 1896 A; 1856 C; 2100 G; 1669 T; 0 other;

Query Match 6.7%; Score 47.4; DB 23; Length 7521;
Best Local Similarity 48.4%; Pred. No. 0.0012;
Matches 132; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

29 GTCTGCTTTTGGGCGCCAGTCTTTTACTGTCCATCGGACCGCAGAAATGCTACTGGTGATTA 88
4494 GTCTTGCACTTGGGCGGCTATGATCTACCGCTCGTCCACAAATGCTTTTGATGA 4553
89 AACAGGAATTAAGCGGAGGACTCATTCGCGTTCTTCTCGTGTGTTTAATTTCTGACG 148
4554 ATCAGGCGATACGTCGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4613
149 TCTTTTGTTCATCGCGGCGGCTGCGGCTGATCTTTTGTCCATGCGCGGCGGATCG 208
4614 TGGTCTGCTGATTTGGCGCGGATTTTGGTGCAGCGGCTTATGATGATGATGATGATGATGAT 4673
209 TGCTCGATATTATGCGTGGGCTGCGATCGCTTACCTGTTATGTTTGGCTCATGGCAG 268
4674 TGCTGGCGCTGCTACCTGGGCGGCTGAACTTCTTCTGTGCAAAACAGACTACCGGG 4733
269 CGAAGAGCGCATGACAAACAGGTGGAGCGC 301
4734 GGAAGAGAGAGCGAAGAGGAGGAGGAGAGC 4766

RESULT 12

AAS94251
ID AAS94251 standard; cDNA; 7521 BP.

AC AAS94251;

DT 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #30055.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG30064.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 1; SEQ ID No 30055; 103pp; English.

PS. The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 7521 BP; 1896 A; 1856 C; 2100 G; 1669 T; 0 other;

Query Match 6.7%; Score 47.4; DB 23; Length 7521;

Best Local Similarity 48.4%; Pred. No. 0.0012;

Matches 132; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 29 GTCTGCTTTTGGGCGGCGATCTTTTACTGTCCATCGGACCGCAGAAATGCTACTGGTGATTA 88

Db 4494 GTCTTGCACTTGGGCGGCTATGATCTACCGCTCGTCCACAAATGCTTTTGATGA 4553

QY 89 AACAGGAATTAAGCGGAGGACTCATTCGCGTTCTTCTCGTGTGTTTAATTTCTGACG 148

Db 4554 ATCAGGCGATACGTCGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4613

QY 149 TCTTTTGTTCATCGCGGCGGCTGCGGCTGATCTTTTGTCCATGCGCGGCGGATCG 208

Db 4614 TGGTCTGCTGATTTGGCGCGGATTTTGGTGCAGCGGCTTATGATGATGATGATGATGAT 4673

QY 209 TGCTCGATATTATGCGTGGGCTGCGATCGCTTACCTGTTATGTTTGGCTCATGGCAG 268

Db 4674 TGCTGGCGCTGCTACCTGGGCGGCTGAACTTCTTCTGTGCAAAACAGACTACCGGG 4733

QY 269 CGAAGAGCGCATGACAAACAGGTGGAGCGC 301

Db 4734 GGAAGAGAGAGCGAAGAGGAGGAGGAGAGC 4766

RESULT 13

AAF71777/c

ID AAF71777 standard; DNA; 993 BP.

AC AAF71777;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
XX fine chemical production; microorganism; organic acid; nucleoside;
XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

PN WO200100843-A2.
XX 04-JAN-2001.
XX 23-JUN-2000; 2000WO-IB00923.
XX 25-JUN-1999; 99US-0141031.
PR 01-JUL-1999; 99DE-1030476.
PR 02-JUL-1999; 99US-0142101.
PR 08-JUL-1999; 99DE-1031415.
PR 08-JUL-1999; 99DE-1031418.
PR 08-JUL-1999; 99DE-1031419.
PR 08-JUL-1999; 99DE-1031420.
PR 08-JUL-1999; 99DE-1031424.
PR 08-JUL-1999; 99DE-1031428.
PR 08-JUL-1999; 99DE-1031434.
PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 08-JUL-1999; 99DE-1032125.
PR 08-JUL-1999; 99DE-1032126.
PR 08-JUL-1999; 99DE-1032130.
PR 08-JUL-1999; 99DE-1032186.
PR 08-JUL-1999; 99DE-1032206.
PR 08-JUL-1999; 99DE-1032227.
PR 08-JUL-1999; 99DE-1032228.
PR 08-JUL-1999; 99DE-1032229.
PR 08-JUL-1999; 99DE-1032230.
PR 08-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 12-AUG-1999; 99US-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-WAR-2000; 2000US-0187970.
XX (BADI) BASF AG.
PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI Lee H, Hwang B;
XX WPI; 2001-137957/14.
DR P-PSDB; AAB79658.
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
microorganisms, including organic acids, nonproteinogenic amino acids,
and purine and pyrimidine bases -
Claim 3; Page 226-228; 1737pp; English.
CC AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;
SQ Query Match 5.5%; Score 39; DB 22; Length 993;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAATCTTCATTACAGGTCGCTTTTG 39
DB 39 ATGGTGATCATGGAATCTTCATTACAGGTCGCTTTTG 1
RESULT 14
AAS96096/c
ID AAS96096 standard; DNA; 993 BP.
XX AC AAS96096;
XX DT 26-FEB-2002 (first entry)
XX DE C. glutamicum gene #21 encoding metabolic pathway protein.
XX KW Methionine biosynthesis pathway; lysine biosynthesis pathway;
XX KW methionine biosynthesis pathway; large-scale production of fine chemical;
XX KW Corynebacterium diphtheriae; diphtheria; ds.
XX OS Corynebacterium glutamicum.
XX WO200166573-A2.
XX 13-SEP-2001.
XX 22-DEC-2000; 2000WO-IB02035.
XX 09-MAR-2000; 2000US-187970P.
XX 23-JUN-2000; 2000US-0606740.
XX (BADI) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
PI Lee H, Hwang B;
XX WPI; 2001-582269/65.
DR P-PSDB; AAB71886.
XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
PT glutamicum, useful for producing methionine and lysine in
PT Corynebacterium and Brevibacterium -
XX Disclosure; Page 207-208; 316pp; English.
XX The present invention relates to the isolation of novel Corynebacterium
CC glutamicum genes encoding metabolic pathway (MP) proteins
CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention
CC include enzymes involved in the lysine and methionine biosynthetic
CC pathways. The polynucleotide sequences of the invention can be used
CC for the large-scale production and/or modulation of expression of
CC fine chemicals such as lysine and methionine. The sequences of the
CC invention may be used to identify C. glutamicum and related organisms
CC e.g. C. diphtheriae in a subject to detect diphtheria.

CC AAS96073-AA96132 represent C. glutamicum genes encoding the novel
CC metabolic pathway proteins of the invention.

XX
SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match 5.5%; Score 39; DB 23; Length 993;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAAATCTTCATTACAGGCTGCTGTTTG 39

DB 39 ATGGTGATCATGGAAATCTTCATTACAGGCTGCTGTTTG 1

RESULT 15

AAI99683/c

ID AAI99683 standard; DNA; 4403765 BP.

AC AAI99683;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where M. tuberculosis strains CDC
PT 1551 and H37Rv differ

PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 5.3%; Score 37.6; DB 22; Length 4403765;

Best Local Similarity 47.5%; Pred. No. 23;

Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 430 AAGCCCATGTTGATCGATCGCTGACCTGTTGATCCGGAATGCGTATTGGACGG 489

DB 720443 AAGGACCTGCGCTGCGCTCGAGCGGGCGGCGGAGCTGCCCGGCCCGGTCGGT 720384

QY 490 TTTGTGTTTATCGGCGCGCTCGCGCGCAATACGGCGACACCGGACGGTGGATTTCGCC 549
DB 720383 GGGCTGCTGTTTACCTGCAACGGGCGCGGACGACGATGTTTCGGGGTCAACGACACGAC 720324
QY 550 GCTGGCGCGTTTCGCGGCAAGCCCTGATCTCGTTTCCCGCTGGTGGGTTTCGGCGCGCAGCAGCA 609
DB 720323 GCGTCGACGATCGAGGACCTGCTGGGCGGGATTCGCTGCGCGGTTTCTTCGCGCGCGG 720264
QY 610 TTGTACGCGCGCTGTTCAGCCGCCCAAGGTGTGGCGCTGGATCAACGTCGTCGTGGC 665
DB 720263 GAGATCGCGCGCTGCGGCGCCACACGCGTTTACCGCGTCGATGGC 720208

Search completed: April 26, 2003, 22:40:54

Job time : 7840.46 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 80.2974 Seconds
(without alignments)
9634.930 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1016_1726

Perfect score: 711

Sequence: 1 atggtgatcatggaatctt.....aactgatgttgagggttag 711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	822	9	US-09-746-660A-51
C 2	711	100.0	3309400	9	US-09-738-626-1
3	708	99.6	708	9	US-09-738-626-1455
4	57	8.0	597	9	US-09-894-844-11
C 5	39	5.5	993	9	US-09-746-660A-47
C 6	37.2	5.2	424	10	US-09-960-352-12392
7	36	5.1	1068	9	US-09-738-626-439
8	36	5.1	3309400	9	US-09-738-626-1
9	35.4	5.0	671	9	US-10-184-644-346
10	35.4	5.0	671	9	US-10-184-634-346
11	35	4.9	5828	9	US-09-430-029-1
12	34.8	4.9	7158	10	US-09-974-300-2171
C 13	34.4	4.8	4372	9	US-09-796-679-4
C 14	34	4.8	2048	10	US-09-925-300-760
C 15	33.8	4.8	2930	10	US-09-745-763-198
C 16	33.8	4.8	88421	9	US-09-976-059-1
17	33.6	4.7	1075	10	US-09-864-761-19241
C 18	33.6	4.7	1403	10	US-09-864-761-2513
C 19	33.6	4.7	1635	10	US-09-864-761-20241

Sequence 3471, Ap
Sequence 36145, A
Sequence 582, App
Sequence 20699, A
Sequence 13829, A
Sequence 20282, A
Sequence 31047, A
Sequence 37138, A
Sequence 2, Appli
Sequence 3, Appli
Sequence 15, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 1545, Ap
Sequence 2257, Ap
Sequence 332, App
Sequence 332, App
Sequence 110, App
Sequence 4, Appli
Sequence 1, Appli
Sequence 27219, A
Sequence 112, App
Sequence 5, Appli
Sequence 5, Appli
Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-746-660A-51
; Sequence 51, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING DATE: 2000-06-23
; PRIOR FILING DATE: 1999-06-25
; PRIOR FILING DATE: 1999-07-02
; PRIOR FILING DATE: 1999-08-12
; PRIOR FILING DATE: 2000-03-09
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patent In Vers. 2.0
; SEQ ID NO 51
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(799)
; OTHER INFORMATION: RXA01394

US-09-746-660A-51

Query Match 100.0%; Score 711; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 7.9e-232;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 60
DB 92 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 151
QY 61 ATCGGACCGAGATGACTGCTGATTAACAAGCAATTAAGCCGAGAGACTCATTTGCG 120
DB 152 ATCGGACCGAGATGACTGCTGATTAACAAGCAATTAAGCCGAGAGACTCATTTGCG 211
QY 121 GTTCTTCTCGTGTTTAAATTTCTGAGCTCTTTTGTTCATCGCCGACCTTGGGCGTT 180
DB 212 GTTCTTCTCGTGTTTAAATTTCTGAGCTCTTTTGTTCATCGCCGACCTTGGGCGTT 271
QY 181 GATCTTTTGTCCAATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 240
DB 272 GATCTTTTGTCCAATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 331
QY 241 TACCTGTTATGTTTCCGCTCATGCGAGCGAAAGACGCGCATGACAAACAAGTGGAAAGCG 300
DB 332 TACCTGTTATGTTTCCGCTCATGCGAGCGAAAGACGCGCATGACAAACAAGTGGAAAGCG 391
QY 301 CCACAGATCATTTGAAGAAACAGAACCGTGCCTCCGATGACACGCTTTGGGCGTTGCG 360
DB 392 CCACAGATCATTTGAAGAAACAGAACCGTGCCTCCGATGACACGCTTTGGGCGTTGCG 451
QY 361 GCGGTGGCCACTGACACGCGCAACCGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 420
DB 452 GCGGTGGCCACTGACACGCGCAACCGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 511
QY 421 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTTTGAACCCGATGCGTAT 480
DB 512 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTTTGAACCCGATGCGTAT 571
QY 481 TTGACGCGTTTGTGTTTATCGGCGGCGTCCGCGCGCAATACGCGGACACCGGACGTTGG 540
DB 572 TTGACGCGTTTGTGTTTATCGGCGGCGTCCGCGCGCAATACGCGGACACCGGACGTTGG 631
QY 541 ATTTTTCGCGCTGGCGGTTCCGCGGAGCTGATCTGTTCCGCTGGGTGGGTGGGTGGGTGG 600
DB 632 ATTTTTCGCGCTGGCGGTTCCGCGGAGCTGATCTGTTCCGCTGGGTGGGTGGGTGGGTGG 691
QY 601 GCAGCAGCATTTGTCAACGCGCGCTGTCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660
DB 692 GCAGCAGCATTTGTCAACGCGCGCTGTCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 751
QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 711
DB 752 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 802

RESULT 2

US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASAO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738.626

; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 100.0%; Score 711; DB 9; Length 3309400;
Best Local Similarity 100.0%; Pred. No. 6.4e-230;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 60
DB 1328953 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 1328894
QY 61 ATCGGACCGAGATGCTACTGCTGATTAACAAGCAATTAAGCCGAGAGACTCATTTGCG 120
DB 1328893 ATCGGACCGAGATGCTACTGCTGATTAACAAGCAATTAAGCCGAGAGACTCATTTGCG 1328834
QY 121 GTTCTTCTCGTGTTTAAATTTCTGAGCTCTTTTGTTCATCGCCGACCTTGGGCGTT 180
DB 1328833 GTTCTTCTCGTGTTTAAATTTCTGAGCTCTTTTGTTCATCGCCGACCTTGGGCGTT 1328774
QY 181 GATCTTTTGTCCAATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 240
DB 1328773 GATCTTTTGTCCAATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 1328714
QY 241 TACCTGTTATGTTTCCGCTCATGCGAGCGAAAGCCCATGACAAACAAGTGGAAAGCG 300
DB 1328713 TACCTGTTATGTTTCCGCTCATGCGAGCGAAAGCCCATGACAAACAAGTGGAAAGCG 1328654
QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCTCCGATGACACGCTTTGGGCGTTGCG 360
DB 1328653 CCACAGATCATTTGAAGAAACAGAACCAACCGTGCCTCCGATGACACGCTTTGGGCGTTGCG 1328594
QY 361 GCGGTGGCCACTGACACGCGCAACCGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 420
DB 1328593 GCGGTGGCCACTGACACGCGCAACCGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 1328534
QY 421 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 480
DB 1328533 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 1328474
QY 481 TTGGACGCGTTTGTGTTTATCGGCGCGCTGCGCGCGCAATACGCGGACACCGGACGTTGG 540
DB 1328473 TTGGACGCGTTTGTGTTTATCGGCGCGCTGCGCGCGCAATACGCGGACACCGGACGTTGG 1328414
QY 541 ATTTTTCGCGCTGGCGGTTTCCGCGCAAGCTGATCTGTTCCGCTGGGTGGGTGGGTGGGTGG 600
DB 1328413 ATTTTTCGCGCTGGCGGTTTCCGCGCAAGCTGATCTGTTCCGCTGGGTGGGTGGGTGG 1328354
QY 601 GCAGCAGCATTTGTCAACGCGCTGTCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660
DB 1328353 GCAGCAGCATTTGTCAACGCGCTGTCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 1328294
QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 711
DB 1328293 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 1328243

RESULT 3
US-09-738-626-3455
; Sequence 3455, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENO, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 3455
LENGTH: 708
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3455

Query Match 99.6%; Score 708; DB 9; Length 708;
Best Local Similarity 100.0%; Pred. No. 7.7e-231;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGATCATGGAATCTTATTACAGTCTCTTTGGGGGCGAGTCTTTTACTGTCC 60
DB 1 ATGTGATCATGGAATCTTATTACAGTCTCTTTGGGGGCGAGTCTTTTACTGTCC 60

QY 61 ATCGGACGCGAGATGACTGCTGATTAACAAGGAATTAAGCGCGAAGGACTCATTTGCG 120
DB 61 ATCGGACGCGAGATGACTGCTGATTAACAAGGAATTAAGCGCGAAGGACTCATTTGCG 120

QY 121 GTTCTTCTCGTGTCTTAATTTCTGAGTCTTTTGTTCATCGCGCGGACCTTGGGGGTT 180
DB 121 GTTCTTCTCGTGTCTTAATTTCTGAGTCTTTTGTTCATCGCGCGGACCTTGGGGGTT 180

QY 181 GATCTTTTGTCCATGCGCGCGCATCGTCTCGATATTATGCTGCGGTGGCGATCGCT 240
DB 181 GATCTTTTGTCCATGCGCGCGCATCGTCTCGATATTATGCTGCGGTGGCGATCGCT 240

QY 241 TACTGTATTGTTGGCTCATGCGCGAAGAACGCGCATGACAAACAAGGTGGAAGCG 300
DB 241 TACTGTATTGTTGGCTCATGCGCGAAGAACGCGCATGACAAACAAGGTGGAAGCG 300

QY 301 CCACAGATCATTTGAAGAACAGAACCAACGCTGCGCATGACAGCCTTTGGGGGTTGCG 360
DB 301 CCACAGATCATTTGAAGAACAGAACCAACGCTGCGCATGACAGCCTTTGGGGGTTGCG 360

QY 361 GCGGTGGCCACTGACGCGCAACGCGTGGGTGGAGGTGAGCGTGCATAGACACGCG 420
DB 361 GCGGTGGCCACTGACGCGCAACGCGTGGGTGGAGGTGAGCGTGCATAGACACGCG 420

QY 421 GTTTGGGTAAAGCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGCAATGCGTAT 480
DB 421 GTTTGGGTAAAGCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGCAATGCGTAT 480

QY 481 TTGACCGGTTTGTATTATCGCGGGGTGCGCGCAATACGCGGACACCGGACGGTGG 540
DB 481 TTGACCGGTTTGTATTATCGCGGGGTGCGCGCAATACGCGGACACCGGACGGTGG 540

QY 541 ATTTTCGCGCTGCGGGTTCGCGGCAAGCTGATCTGTTCCGCTGGTGGTTCGCG 600
DB 541 ATTTTCGCGCTGCGGGTTCGCGGCAAGCTGATCTGTTCCGCTGGTGGTTCGCG 600

QY 601 GCACGACATTTGTACGCGCGCTGCTCCAGCCCCCAAGGTGTGGGCTGGATCAACGTCGTC 660

DB 601 GCACGACATTTGTACGCGCGCTGCTCCAGCCCCCAAGGTGTGGGCTGGATCAACGTCGTC 660

QY 661 GTGGCAGTTGTGATGACCGGCAATGGCCATCAAACTGATGTTGATGGT 708

DB 661 GTGGCAGTTGTGATGACCGGCAATGGCCATCAAACTGATGTTGATGGT 708

RESULT 4

US-09-894-844-11
Sequence 11, Application US/09894844
Patent No. US20020176873A1
GENERAL INFORMATION:
APPLICANT: Behr, Marcel
APPLICANT: Small, Peter
APPLICANT: Schoonik, Gary
APPLICANT: Wilson, Michael A.
TITLE OF INVENTION: Molecular Differences Between Species of
FILE REFERENCE: the M. Tuberculosis Complex
FILE REFERENCE: STAN102CON
CURRENT APPLICATION NUMBER: US/09/894,844
CURRENT FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: 09/318,191
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/097,936
PRIOR FILING DATE: 1998-08-25
NUMBER OF SEQ ID NOS: 137
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 597
TYPE: DNA
ORGANISM: Mycobacteria tuberculosis
US-09-894-844-11

Query Match 8.0%; Score 57; DB 9; Length 597;
Best Local Similarity 53.1%; Pred. No. 4.9e-09;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATCGTCTGACCTGTTGATGACCGGATGTTGGAGCGT 490
DB 326 AGTCTCTGTTGACCTTGGCGGCAATTCACGTTCTCTCAACCCACACGCTCTACCTCGACACCG 385

QY 491 TTGTGTTTATCGCGCGCTCGCGGCAATACGGCGACACCGGCGATTTTCGCGG 550
DB 386 TCGTGTGTTGAGCGGCTGGCGACAGCAGCA--CCAGCGCTGGCTGTTTCGCGC 442

QY 551 CTGGCGGTTTCGCGCAAGCCTGATCTGTTCCCGTGGTGGGTTTCGCGCGAGCAGCAT 610
DB 443 TCGCGCGGTCACAGCAGTGGGTATGTTCCGACCCCTCGGGTTCGAGCGCGCGT 502

QY 611 TGTACGCGCGCTGTCAGCCCCCAAGGTGTGGGCTGGATCAACGTCGTTGGGAGTTG 670
DB 503 TCGCGCGGCTGTTACCAACCCCGGCTCGTGAGAAATCTCGACGCGCTGATCGCGGTCA 562

QY 671 TGATGACCGCATTTGGCCATCAAACTGATTTGA 703

DB 563 TGATGTTGCGTGGGAATCTCGCTGACCGTGA 595

RESULT 5

US-09-746-660A-47/c
Sequence 47, Application US/09746660A
Publication No. US20030049804A1
GENERAL INFORMATION:
APPLICANT: Pompejus, Markus
APPLICANT: Kroger, Burkhard
APPLICANT: Schroder, Hartwig
APPLICANT: Zelder, Oskar
APPLICANT: Habernauer, Gregor
APPLICANT: Kim, Jun-Won
APPLICANT: Lee, Heung-Schick
APPLICANT: Hwang, Byung-Joon
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING

;; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
;; FILE REFERENCE: BGI-121CP2
;; CURRENT APPLICATION NUMBER: US/09/746,660A
;; CURRENT FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 09/606740
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: 09/603124
;; PRIOR FILING DATE: 2000-06-23
;; PRIOR APPLICATION NUMBER: 60/141031
;; PRIOR FILING DATE: 1999-06-25
;; PRIOR APPLICATION NUMBER: 60/142101
;; PRIOR FILING DATE: 1999-07-02
;; PRIOR APPLICATION NUMBER: 60/148613
;; PRIOR FILING DATE: 1999-08-12
;; PRIOR APPLICATION NUMBER: 60/187970
;; PRIOR FILING DATE: 2000-03-09
;; PRIOR APPLICATION NUMBER: DE 19931420.9
;; PRIOR FILING DATE: 1999-07-08
;; NUMBER OF SEQ ID NOS: 125
;; SOFTWARE: PatentIn Vers. 2.0
;; SEQ ID NO 47
;; LENGTH: 993
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (101)..(970)
;; OTHER INFORMATION: RXA01393
US-09-746-660A-47

Query Match 5.5%; Score 39; DB 9; Length 993;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGTGATCATGGAATCTTATTACAGGCTGCTTTTG 39
Db 39 ATGGTGATCATGGAATCTTATTACAGGCTGCTTTTG 1

RESULT 6

US-09-960-352-12392/c
;; Sequence 12392, Application US/09960352
;; Patent No. US20020137139A1
;; GENERAL INFORMATION:
;; APPLICANT: Warren, Wesley C.
;; APPLICANT: Tao, Nengboing
;; APPLICANT: Byatt, John C.
;; APPLICANT: Mathialagan, Nagappan
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;; FILE REFERENCE: 16511.006/37-21(10298)C
;; CURRENT APPLICATION NUMBER: US/09/960,352
;; CURRENT FILING DATE: 2001-09-24
;; NUMBER OF SEQ ID NOS: 15112
;; SEQ ID NO 12392
;; LENGTH: 424
;; TYPE: DNA
;; ORGANISM: Bos taurus
;; OTHER INFORMATION: Clone ID: 53-LIB3058-002-Q1-K1-F2
US-09-960-352-12392

Query Match 5.2%; Score 37.2; DB 10; Length 424;
Best Local Similarity 47.8%; Pred. No. 0.023;
Matches 108; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 349 TTGGCGCGTTCGGCGTGGCCACTGACACGCGCAACCGGGTGGGGTGGAGGTGAGCGTC 408
Db 301 TTGGCGTGGACGCGCGAGAGTCCGAGATGCGCCCTGGAGAGCTTGGCGGTTATGTC 242
Qy 409 GATAAGCAGCGGTTTGGGTAAGCCCATTTGATGGCAATCGTGTGACCTGTTGAAC 468
Db 241 CTCCTGTTAGACGAGGTGGGAGCCGGAAGTGGAGCTGACGAGGTGCTGGCGAGGTGATC 182

Qy 469 CCGAATGCGTATTGTGACGCGTGTGTTTATCGCGCGCGTTCGGCGCAATACAGCGAC 528
Db 181 TCGAAGCGCGCGTGGCTTCGTCGAAGCTGTAGAGGTACCGAGTCCGGGATGAGCCCC 122
Qy 529 ACCGAGCGTGGATTTTCGCGCGTGGCGGCTTCGGCGCAAGCCTGA 574
Db 121 TTCGGGAGGCGGTACTTTGGGACAGAGTGTTCGCGCGCGCGCTGA 76

RESULT 7

US-09-738-626-439
;; Sequence 439, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 439
;; LENGTH: 1068
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-439

Query Match 5.1%; Score 36; DB 9; Length 1068;
Best Local Similarity 53.6%; Pred. No. 0.095;
Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 551 CTGGCGCGTTCGGCGCAAGCTGATCTGTTCCCGTGGTGGTTCGGCGCAGCAGCAT 610
Db 542 CCGGCGCGTGCACCTGTTCTTCTATCATCGCGCCGACGAGCTCCCGGAAGAAATTA 601
Qy 611 TGTACGCGCGTGTCCAGCCCCCAAGGTGTGGGCTGGATCAACGTGCTGGCGAGTTG 670
Db 602 TTTTCTGGCAGATGGTTCCTCAACGGCTCGCAATGGCGGACGTCGACGTGTTG 661
Qy 671 TGATGACCGCATTTGCCATC 690
Db 662 TGATGTCGAATCGGCCTC 681

RESULT 8

US-09-738-626-1
;; Sequence 1, Application US/09738626
;; Publication No. US20020197605A1
;; GENERAL INFORMATION:
;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; PRIOR FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent in ver. 3.0
 ; SEQ ID NO 1
 ; LENGTH: 3309400
 ; TYPE: DNA
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1

Query Match 5.1%; Score 36; DB 9; Length 3309400;
 Best Local Similarity 53.6%; Pred. No. 6.6;
 Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 551 CTGGCGCGTTCGCGCAAGCTGATCTGTTCCCGCTGTTGCGGCGAGCAGCAT 610
 Db 414184 CGGCGCGCTGACCTCGTTCTTCTATCTACATCGCGCGAGCTCCCGCGAAGAAATTA 414243
 QY 611 TGTACGCGCGCTGTCACGCCCAAGGTGTGGCGTGGATCAAGCTGCTGCGCAGTTG 670
 Db 414244 TTTTCTGCAGATGGTTCCCTCAACGGCTCGCAATGGCGCAGCTCAACGTGGTGTG 414303
 QY 671 TGATGACCGCATGGCCATC 690
 Db 414304 TGATGTCACAATCGGCTC 414323

RESULT 9
 US-10-184-644-346
 ; Sequence 346, Application US/10184644
 ; Publication No. US20030044930A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430RIC227
 ; CURRENT APPLICATION NUMBER: US/10/184,644
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 346
 ; LENGTH: 671
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-184-644-346

Query Match 5.0%; Score 35.4; DB 9; Length 671;
 Best Local Similarity 6.3%; Pred. No. 0.12;
 Matches 29; Conservative 166; Mismatches 265; Indels 0; Gaps 0;
 QY 108 AGGACTCATTGCGGTCTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGG 167
 Db 32 AVKPPNKPPIFFPGTHETAPLPGKDLFPYDKCKDKYKPKRKGPNGLWEIQNNPHAS 91

QY 168 CACCTTGGCGGTGATCTTTTGTCCAAATGCGCGCCGATCGTCTGATATTATGCGCTG 227
 Db 92 YSAPPVSSSDSEAPANPADGSDADDDDRGVMATVATATAADRMESSDSKSSD 151
 QY 228 GGGTGGCATCGTTACCTGTTATGTTTGGTCCGTCATGGCAGCAAGAACCCATGACAAA 287
 Db 152 NSGLKRTKTPALKMSVSKARKASDLDAQSVSPSEENSSSESEKTSQDQFTPEKAA 211
 QY 288 CAAGGTGAAGCGCCACAGATCATTGAGAAACAGAACCAACCGTCCCGCATCACGCC 347
 Db 212 VRAPRRGLGRKKKKAPSASDSKADSGAKPEPVMARSASSSSSSSDSDSVX 271
 QY 348 TTTGGCGGTTCGCGGTGCGCACTGACACGCGCAACCGGTCGCGTGAGGTGAGCGT 407
 Db 272 KPPGRKPAEKPLPKPRGRKPKPERPPSSSSSDSDSDVDRISWKRRDRREARR 331
 QY 408 CGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGAA 467
 Db 332 RREQEELRLREQEKEKERRERADRGAEKRGSGSSGDELREDDPEPVKGRKGRGR 391
 QY 468 CCCGAATCGTATTGACGCGGTTTGTGTTTATCGGCGCGTGGCGCGCAATACGCGA 527
 Db 392 GPSSSDSEPEALEERAKSKAKPOSSSTEPARKPEKVRPEEKQAKPVKVERTR 451
 QY 528 CACGACGCGGTGATTTTCGCCGCTGCGCGTTCGCGGCAAG 569
 Db 452 KRSEGFMDRKEKKEPSVEEKLQKLHSEIKALKVDSVDV 493

RESULT 10
 US-10-184-634-346
 ; Sequence 346, Application US/10184634
 ; Publication No. US2003006864A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430RIC217
 ; CURRENT APPLICATION NUMBER: US/10/184,634
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 346
 ; LENGTH: 671
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-184-634-346

Query Match 5.0%; Score 35.4; DB 9; Length 671;
 Best Local Similarity 6.3%; Pred. No. 0.12;
 Matches 29; Conservative 168; Mismatches 265; Indels 0; Gaps 0;
 QY 108 AGGACTCATTGCGGTCTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGG 167
 Db 32 AVKPPNKPPIFFPGTHETAPLPGKDLFPYDKCKDKYKPKRKGPNGLWEIQNNPHAS 91
 QY 168 CACCTTGGCGGTGATCTTTTGTCCAAATGCGCGCGATCGTCTGATATTATGCGCTG 227
 Db 92 YSAPPVSSSDSEAPANPADGSDADDDDRGVMATVATATAADRMESSDSKSSD 151
 QY 228 GGGTGGCATCGTTACCTGTTATGTTTGGTCCGTCATGGCAGCAAGAACCCATGACAAA 287
 Db 152 NSGLKRTKTPALKMSVSKARKASDLDAQSVSPSEENSSSESEKTSQDQFTPEKAA 211

QY 288 CAAGGTGAAGCCGACAGATCATTTGAAGAAACAGAACCAACCGTGCCTGATGACACGCC 347
Db 212 VRAPRRGFLGKRRKKKAPSDSKADSDGAKPEPVAMARASSSSSSSSSSSSDSVK 271
QY 348 TTTGGCGGTTTCGGCGGTGGCCACTGACACCGCAACCGGTGCGGTGGAGGTGACGGT 407
Db 272 KPPGRKPAEKPLPKPRGRPKPRPPSSSSSSSDSDSDSDSDSDSDSDSDSDSDSDSD 331
QY 408 CGATAAGCAGCGGTTTGGGTAAAGCCCATCTTGATGCAATCTGCTGACCTGTTGAA 467
Db 332 RREQEELRLRLREQEKEKERRERADGEAEAGSGSSGDELREDEPVKGRKGRGR 391
QY 468 CCGAATCGGTATTTGACGCGTTTGTGTTATCGGCGCGTCTCGCGCGCAATACGCGGA 527
Db 392 GPPSSSDSEPAELERAEAKSAKPPQSSSTEPARKPGQKEKRVPRBKQAKPVKVBTR 451
QY 528 CACCGGCGGTGATTTTCGCGCGTGGCGCTTCGCGGCAAG 569
Db 452 KRSEGFMDRKVEKKKPSVEKLQKLHSEIKFALKVDSVDV 493

RESULT 11

US-09-430-029-1
; Sequence 1, Application US/09430029
; Patent No. US20020168738A1
; GENERAL INFORMATION:
; APPLICANT: Yano, Tetsuya; No. US20020168738A1, tsuyoshi; Imamura, Takeshi; Canon K
; TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene,
; TITLE OF INVENTION: Recombinant Plasmid, Transformed Microorganism,
; TITLE OF INVENTION: Method for Degrading Chlorinated Aliphatic Hydrocarbon
; TITLE OF INVENTION: Compounds and Aromatic Compounds, and
; TITLE OF INVENTION: Method for Environmental Remediation
; FILE REFERENCE: CPO13982US
; CURRENT APPLICATION NUMBER: US/09/430,029
; CURRENT FILING DATE: 1999-10-29
; EARLIER APPLICATION NUMBER: JP P1998-310801
; EARLIER FILING DATE: 1998-10-30
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 5828
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234)..(443)
; OTHER INFORMATION: tomk
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (463)..(1455)
; OTHER INFORMATION: tomL
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1495)..(1761)
; OTHER INFORMATION: tomM
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1803)..(3350)
; OTHER INFORMATION: tomN
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3428)..(3781)
; OTHER INFORMATION: tomO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3810)..(4871)
; OTHER INFORMATION: tomP
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4876)..(5229)
; OTHER INFORMATION: tomQ
US-09-430-029-1

Query Match 4.9%; Score 35; DB 9; Length 5828;
Best Local Similarity 50.3%; Pred. No. 0.51;
Matches 86; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 322 GAACCAACCGTGCCTGATGACACCGCTTTGGCGGTTTCGGCGGTGCGCACTGACACGCC 381
Db 5499 GAAGCGTACGGGATCCGACCGAGATGCTGCCGAAGCGCGCTGCCGCGGTGCGCCGC 5558
QY 382 AACCGGTCGGGTGAGGTGAGCGTGAAGCAACGCGGTTTGGGTAAAGCCCATGTTG 441
Db 5559 CAACTCGGTTCTGCTGCCGAGCGCCATGAATCGCGCTGTTCCGGAAGAGGCGCTG 5618
QY 442 ATGGCAATCGTGTGACCTGTTGAACCGCAATGCGTATTTGACGCGTTT 492
Db 5619 GTGGCACCAGCGTCCGCTCGTGAACCCCGATCCGTGCGCCGACGACATT 5669

RESULT 12

US-09-974-300-2171
; Sequence 2171, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2171
; LENGTH: 7158
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2171

Query Match 4.9%; Score 34.8; DB 10; Length 7158;
Best Local Similarity 54.8%; Pred. No. 0.66;
Matches 69; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 450 CGTGTGACCTGGTTGAACCCGGAATGCTATTTGGACGCGTTTGTGTTATCGCGCGCT 509
Db 5650 CGGGTTACTTGGCAGACCCGATTGACAGACGCGTTTGTGCTGATCCGTTGGT 5709
QY 510 CGGCGCGCAATACGCGACACCGGACGCGTGGATTTTCGCGCGTGGCGGTTTCGCGCAAG 569
Db 5710 CCGCGGGAACCGGATGTACCGCACCGCGATCTTGGCCGAGCGCGCTCT 5769
QY 570 CCTGAT 575
Db 5770 CTGGAT 5775

RESULT 13

US-09-796-679-4/c
; Sequence 4, Application US/09796679
; Publication No. US20030013076A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Anthony J
; APPLICANT: Lyttle, David J
; APPLICANT: University of Otago
; TITLE OF INVENTION: Parapoxvirus vectors
; FILE REFERENCE: 23607 MRB
; CURRENT APPLICATION NUMBER: US/09/796,679
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/155,421
; PRIOR FILING DATE: 1998-09-29

PRIOR APPLICATION NUMBER: PCT/NZ97/00040
PRIOR FILING DATE: 1997-03-27
PRIOR APPLICATION NUMBER: NZ 286284
PRIOR FILING DATE: 1996-03-29
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 4372
TYPE: DNA
ORGANISM: Orf virus strain NZ-2
US-09-796-679-4

Query Match 4.8%; Score 34.4; DB 9; Length 4372;
Best Local Similarity 43.8%; Pred. No. 0.7;
Matches 149; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

Qy 364 GTGGCCACTGACACGCGGAGCGGTGGGTGGAGTGGAGTGGATGAAGCAGCGGTT 423
Db 1493 GAGCGCGCGGTGGGGGAGGGGTGGCTTCGAGTCCGCCCTCGGAGCAGCAGATC 1434

Qy 424 TGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGGTTGAACCCCAATGCGTATTG 483
Db 1433 CGCTCGCGGTGGATGCGGTGACGTGACTTTCGCCAAGATCAAGAGTTCTGTG 1374

Qy 484 GACGGTTTGTGTTATCGCGGGGTGCGCGGCAATACCGCGACACCGGACGGTGGATT 543
Db 1373 GAGGAGTTCGTGTACTTTGAGACGCGCATGCGCTACTCGCGCATCGGCGATCAATC 1314

Qy 544 TTCGCGGTGGCGGTTCGCGGCAAGCTGATCTGTTCCGCTGGTGGTTCGGCGCA 603
Db 1313 CCGGACTGAACCGACGCGTTCGCGCGCGCGCGCGGTGGTGGTCCCGGACCAAC 1254

Qy 604 GCAGCATTTGCACGCCCTGTTCAGCCCCAAGGTGGCGCTGATCAACGTGCTGTG 663
Db 1253 AAGTCCATCTTCTTCAGCGAGCCCTACAGTACTTCTCGACAGCCGCTTCACTTC 1194

Qy 664 CGAGTTGTGATGACCGCATGCGCATCAAACTGATTTGA 703
Db 1193 AACATCGTGATGCTCTCGACACCATCATGAAGCGCAGA 1154

RESULT 14
US-09-925-300-760/c
Sequence 760, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben,
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 760
LENGTH: 2048
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1957)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (1963)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (2006)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-760

Query Match 4.8%; Score 34; DB 10; Length 2048;
Best Local Similarity 56.1%; Pred. No. 0.64;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 393 GGTGGAGTGCAGCTGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCAATCGT 452
Db 1381 GGTGACCACTGATGACATGTTGGCGTGGCTGCAGCCCTGCTGGGAGCCT 1322

Qy 453 GCTGACCTGTTGAACCGCAATCGTATTGGAGCGGTTTGTGTTATCGGCGG 506
Db 1321 GCTGACCTGAGGAGCGGAAAGTGCATTTGACCGTGTGTTTGTGACAGGCAG 1268

RESULT 15
US-09-745-763-198/c
Sequence 198, Application US/09745763
Patent No. US20020065394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
LaVallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Treacy, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 198:
SEQUENCE CHARACTERISTICS:
LENGTH: 2930 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 198:
US-09-745-763-198

Query Match 4.8%; Score 33.8; DB 10; Length 2930;
Best Local Similarity 49.7%; Pred. No. 0.91;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 415 CAGCGGGTTTGGTAAAGCCCATGTTGATGGCAATCGTGCACCTGTTGAACCCGAAT 474
Db 1652 CGGGGGCTTGGTCCCACTGGGGCCCTGAGGGCGGCTTCCAGAGGGAACACCGGA 1593

Qy 475 GCGTATTGACGCGTTTGTGTTTATCGCGCGCGTGGCGCGCAATACGCGCACACCGGA 534
Db 1592 GCCTTTGGGGCCCTGGGAGCCTTTAGATCCTTTTGGCGCCACGCTCTCCGGGGGACCG 1533

Qy 535 CGGTGATTTTCGCCGCTGGCGCGTTCCGGCAAGCCTGATCTGGTTCCCGCT 587
Db 1532 TGGTCCAATTGGGCTCTCTCACCCGACGGCCAGGTGGTCCAGGCTCCCGCT 1480

Search completed: April 27, 2003, 04:41:53
Job time : 2873.3 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:22:28 ; Search time 40.0589 Seconds
(without alignments)
5443.169 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1016_1726

Perfect score: 711

Sequence: 1 atggtgatcatggaatctt.....aactgatgtgtaggtag 711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

1: /cgn2_6/prodata/2/ina/5A COMB.seq.*

2: /cgn2_6/prodata/2/ina/5B COMB.seq.*

3: /cgn2_6/prodata/2/ina/6A COMB.seq.*

4: /cgn2_6/prodata/2/ina/6B COMB.seq.*

5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*

6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	57	8.0	15239	1	US-08-390-878-17
2	57	8.0	4403765	4	US-09-103-840A-2
3	57	8.0	4411529	4	US-09-103-840A-1
C 4	37.6	5.3	4403765	4	US-09-103-840A-2
C 5	37.6	5.3	4411529	4	US-09-103-840A-1
6	34.6	4.9	6611	1	US-08-402-282-2
7	34.6	4.9	6611	1	US-08-508-004-2
8	34.6	4.9	6611	1	US-08-402-066-2
9	34.6	4.9	6611	1	US-08-402-068-2
10	33.6	4.7	1681	4	US-09-434-288-7
11	33	4.6	4826	4	US-09-192-983-3
12	32.8	4.6	3955	4	US-09-214-278-4
13	32.8	4.6	4315	3	US-08-882-046-3
14	32.8	4.6	4483	3	US-08-611-729A-7
C 15	32.8	4.6	4508	5	PCT-US93-06251-34
16	32.4	4.6	707	4	US-08-998-416-940
17	32.4	4.6	4689	4	US-09-105-537-34
18	32.4	4.6	36778	4	US-09-105-537-5
19	32.4	4.6	38506	3	US-09-320-878-19
20	32.2	4.5	1404	4	US-08-979-608A-10
21	32	4.5	43280	2	US-08-804-227C-1
22	31.8	4.5	1050	4	US-09-161-241-77
C 23	31.8	4.5	4550	4	US-09-338-907-182
C 24	31.8	4.5	4550	4	US-09-218-207-182
25	31.6	4.4	50937	4	US-09-428-517-1
C 26	31.4	4.4	50937	4	US-09-428-517-1
C 27	31.2	4.4	219	3	US-08-078-271B-9

28	31.2	4.4	1125	2	US-08-997-080-153	Sequence 153, Appl
29	31.2	4.4	1125	2	US-08-997-362-153	Sequence 153, Appl
30	31.2	4.4	1125	4	US-09-095-855-153	Sequence 153, Appl
31	31.2	4.4	1125	4	US-09-324-542-153	Sequence 153, Appl
32	31.2	4.4	1125	4	US-09-205-426-153	Sequence 17, Appl
33	31	4.4	289	4	US-09-007-005-17	Sequence 17, Appl
34	31	4.4	289	4	US-09-244-796-17	Sequence 1, Appl
35	31	4.4	1365	4	US-09-319-892-1	Sequence 1, Appl
36	31	4.4	1917	4	US-09-503-172A-1	Sequence 1, Appl
37	31	4.4	3078	4	US-09-418-817-9	Sequence 9, Appl
38	31	4.4	6217	4	US-09-418-817-1	Sequence 1, Appl
C 39	30.8	4.3	735	3	US-09-003-287-7	Sequence 7, Appl
C 40	30.8	4.3	1865	4	US-09-370-253-5	Sequence 5, Appl
41	30.6	4.3	1437	1	US-08-252-966B-14	Sequence 14, Appl
C 42	30.6	4.3	4161	3	US-08-790-517-1	Sequence 1, Appl
43	30.6	4.3	4161	3	US-08-790-517-19	Sequence 19, Appl
44	30.6	4.3	4212	4	US-09-221-017B-39	Sequence 39, Appl
45	30.6	4.3	8931	3	US-09-028-934-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-390-878-17/c
; Sequence 17, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-17

Query Match 8.0%; Score 57; DB 1; Length 15239;
Best Local Similarity 53.1%; Pred. No. 4.5e-07;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 ACCCATGTTGATGTCATCGTGGTACCTGGTGAACCCGATCGGATTTTGGACGGCT 490

DB 4792 AGGTCTGTGTGCTGCGGCAATTCAGTTCTCAACCCACGCTACCTCGACCG 4733

```
QY 491 TTGTTGTTATCGCGCGCTCGCGCGCAATACGCGACACCGGACGCTGATTTTCGCCG 550
DB 4732 TCCTGTTGCTAGCGCGCTCGCGCAACGAGCACACGCA---CCAGCGTGGCTTTGCCG 4676
QY 551 CTGCGCGGTTTCGGCGCAAGCCTGATCTGTTCCCGTGTGGTTCGGCGCAGCAGCAT 610
DB 4675 TCGGCGCGGTTCACGACGATGCGGTATGTTCCGCCACCTCGGTTTCGGAGCGCGCGGT 4616
QY 611 TGTACGCGCGCTGTTCAGCGCCCAAGGTGGCGCTGGATCAACGTCGTGGCAGTTG 670
DB 4615 TCGCGCGGCTGTTTCAACAACCGCGCTCGTGAGAAATCTCTCGACGCGCTGATCGCGGTCA 4556
QY 671 TGATACCGCATTTGGCCATCAAACTGATGTTGA 703
DB 4555 TGATGTTGCGCTGGGAATCTCGTACCGGTGA 4523
```

RESULT 2

```
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 8.0%; Score 57; DB 4; Length 4403765;
Best Local Similarity 53.1%; Pred. No. 6.9e-06;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGAATGCGTATTTGGACGCGT 490
DB 2227633 AGTCTCTGTGACCTGTGGGCAATTCAGTTCCTCAACCCACAGCTTACCTCGACACCG 2227692
QY 491 TTGTTGTTATCGCGCGCTCGCGCGCAATACGCGACACCGGACGCTGATTTTCGCCG 550
DB 2227693 TCGTGTGTTAGCGCGCTGGCCACGACACAGCA---CCAGCGCTGGCTTTCCGCC 2227749
QY 551 CTGCGCGGTTTCGGCGCAAGCCTGATCTGTTCCCGTGTGGTTCGGCGCAGCAGCAT 610
DB 2227750 TCGCGCGGTTCACGACGATGCGGTATGTTTCGCCACCTCGGTTTCGGAGCGCGCGT 2227809
QY 611 TGTACCGCGCTGTCTCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTGGCAGTTG 670
DB 2227810 TCGCGCGGTGTTTCAACCAACCGCGCTCGTGGAATCTCGACGCGCTGATCGCGGTCA 2227869
QY 671 TGATACCGCATTTGGCCATCAAACTGATGTTGA 703
DB 2227870 TGATGTTGCGCTGGGAATCTCGTACCGGTGA 2227902
```

RESULT 3

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
```

```
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
Query Match 8.0%; Score 57; DB 4; Length 4411529;
Best Local Similarity 53.1%; Pred. No. 6.9e-06;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGAATGCGTATTTGGACGCGT 490
DB 2230334 AGTCTCTGTGACCTGTGGGCAATTCAGTTCCTCAACCCACAGCTTACCTCGACACCG 2230393
QY 491 TTGTTGTTATCGCGCGCTCGCGCGCAATACGCGCAGCAGCGGATGATTTTCGCCG 550
DB 2230394 TCGTGTGTTAGCGCGCTGGCCAAACGACACAGCA---CCAGCGCTGGCTTTCCGCC 2230450
QY 551 CTGCGCGGTTTCGGCGCAAGCCTGATCTGTTCCCGTGTGGTTCGGCGCAGCAGCAT 610
DB 2230451 TCGCGCGGTTCACGACGATGCGGTATGTTTCGCCACCTCGGTTTCGGAGCGCGCGT 2230510
QY 611 TGTACCGCGCTGTCTCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTGGCAGTTG 670
DB 2230511 TCGCGCGGTGTTTCAACCAACCGCGCTCGTGGAATCTCGACGCGCTGATCGCGGTCA 2230570
QY 671 TGATACCGCATTTGGCCATCAAACTGATGTTGA 703
DB 2230571 TGATGTTGCGCTGGGAATCTCGTACCGGTGA 2230603
```

RESULT 4

```
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
Query Match 5.3%; Score 37.6; DB 4; Length 4403765;
Best Local Similarity 47.5%; Pred. No. 4.7;
Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
```



```

Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 359 CGCGGTGGCCACTGACACGGCAACCGGTGGGGTGGAGGTGAGTGCATAGCAGC 418
      |||||
Db 619 CGCGCATCGCTAAGCGCTTCCGAAGCGGGTGGCGGTGAAGTTTACGCACGCGGTGCGC 678
      |||||
QY 419 GGGTTTGGGTAAGCCCATCTTGTATGCAATCGTGTGACCTGTTCAACCCGATGCGT 478
      |||||
Db 679 GGGCTGTGGTGACAGCGGGGGCGCGGAGTTCACGTGCGCCACGATGAGCCCGGGAGT 738
      |||||
QY 479 ATTGACGCGGTTTGTGTTTATCGCGCGCGTGGCGGCGCAATACGGCGA 527
      |||||
Db 739 TCCGCACTGCCCGGCTGCCACCGAGGGCGGCGATCGTGGACGGCA 787
      |||||

```

```

RESULT 7
US-08-508-004-2
; Sequence 2, Application US/08508004
; Patent No. 5582969
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Bever Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,004
; FILING DATE: 27-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,282
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6611 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 390..1538
; OTHER INFORMATION: /function= "coding sequence"
; OTHER INFORMATION: /product= "DNA polymerase"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2107..3132
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: /frame= "frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3138..3359
; OTHER INFORMATION: /function= "potential open reading"

```

```

; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4690..5028
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5028..5375
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5375..5653
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5653..5910
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6078..6491
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; US-08-508-004-2
Query Match 4.9%; Score 34.6; DB 1; Length 6611;
Best Local Similarity 50.3%; Pred. No. 1.7;
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 359 CGCGGTGGCCACTGACACGGCAACCGGTGGGGTGGAGGTGAGTGCATAGCAGC 418
      |||||
Db 619 CGCGCATCGCTAAGCGCTTCCGAAGCGGGTGGCGGTGAAGTTTACGCACGCGGTGCGC 678
      |||||
QY 419 GGGTTTGGGTAAGCCCATCTTGTATGCAATCGTGTGACCTGTTCAACCCGATGCGT 478
      |||||
Db 679 GGGCTGTGGTGACAGCGGGGGCGCGGAGTTCACGTGCGCCACGATGAGCCCGGGAGT 738
      |||||
QY 479 ATTGACGCGGTTTGTGTTTATCGCGCGCGTGGCGGCGCAATACGGCGA 527
      |||||
Db 739 TCCGCACTGCCCGGCTGCCACCGAGGGCGGCGATCGTGGACGGCA 787
      |||||
RESULT 8
US-08-402-066-2
; Sequence 2, Application US/08402066
; Patent No. 5612182
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Bever Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,066
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6611 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 390..1538
; OTHER INFORMATION: /function= "coding sequence"
; OTHER INFORMATION: /product= "DNA polymerase"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2107..3132
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: /frame= "frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3138..3359
; OTHER INFORMATION: /function= "potential open reading"

```



```

:
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
:
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 6078..6491
: OTHER INFORMATION: /function= "potential open reading
: OTHER INFORMATION: frame"
:
US-08-402-068-2

```

	Query Match	4.9%;	Score 34.6;	DB 1;	Length 6611;
	Best Local Similarity	50.3%;	Pred. No. 1.7;		
	Matches 85;	Conservative 0;	Mismatches 84;	Indels 0;	Gaps 0;
Qy	359	CGGCGGTGGCCACTGACACGCGCAACCGGGTGGCGGTGAGCGTGCATGAAGCAGC	418		
Db	619	CGGCATCGCTAAGCGTTGCCGACCGGGTCCCGTGAACTTTACGCACACCGGTGCGC	678		
Qy	419	GGGTGTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCCGGAATGCGT	478		
Db	679	GGGTGTGTGTGACAGCGGGGGCCGCGGAGTTCAAGCTCTCCACAGATGACCCGCGGGAGT	738		
Qy	479	ATTTGGACGCGTTGTGTGTTATTCGGCGCGGTCTGGCGCGCAATACGGCGA	527		
Db	739	TCCGCGCAACTGCCCGGCTGCGCCACCGAGCGGGGATCGTGGACGGCGCA	787		

RESULT 10
 US-09-434-288-7
 ; Sequence 7, Application US/09434288
 ; Patent NO. 6303767
 ; GENERAL INFORMATION:
 ; APPLICANT: Betlach C., Melanie
 ; APPLICANT: McDaniel, Robert
 ; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
 ; TITLE OF INVENTION: CONSTRUCTS THEREFOR
 ; FILE REFERENCE: 30062-20030.00
 ; CURRENT APPLICATION NUMBER: US/09/434,288
 ; CURRENT FILING DATE: 1999-11-05
 ; PRIOR APPLICATION NUMBER: 60/107,093
 ; PRIOR FILING DATE: 1998-11-05
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 1681
 ; TYPE: DNA
 ; ORGANISM: Streptomyces carbonensis
 US-09-434-288-7

	Query Match	4.7%	Score 33.6;	DB 4;	Length 1681;
	Best Local Similarity	49.4%;	Prod. No. 1.8;		
	Matches 87;	Conservative	0;	Mismatches 89;	Indels 0;
	Gaps	0;			
Qy	520	TACGGCGACACGGACGGTGGNTTTTCGCCGTGCGGGTTCGGCGCAGCCTGATCTGG	579		
Db	1493	TACTCCCGGTTCGGCGGGTTCCTCCACGACGGCGGCGATTTCGACGCCGACTTCTTCGGG	1552		
Qy	580	TTCCCGCTGTGTGGGTTTCGGCCGACGACGATTTGTTCAGCCCGCTGTTCAGCCGCCAAGGTG	639		
Db	1553	ATTCGCGCGCGAGGCCCTCGCCATGACCCGCGACGAGCGGCTGTCCCTCACACCGCG	1612		
Qy	640	TGGCGCTGGATCAACGTGTGTGTGGAGTTGTGTGATGACCGCATTTGGCCATCAAACT	695		
Db	1613	TGGGAGGGCGATCGACACGCGGGCATTCACCCGACGAGCTCAAGGGCAGCGGCTT	1668		

RESULT 11
US-09-192-983-3
; Sequence 3, Application US/09192983A
; Patent No. 6242244
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy
; APPLICANT: Barber, Robert
; APPLICANT: Witthuhn, Vernon

```

; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and
;
; FILE OF INVENTION: Remediation
;
; FILE REFERENCE: 960296.95505
;
; CURRENT APPLICATION NUMBER: US/09/192,983A
;
; CURRENT FILING DATE: 1998-11-16
;
; EARLIER APPLICATION NUMBER: 08/919,953
;
; EARLIER FILING DATE: 1997-08-29
;
; EARLIER APPLICATION NUMBER: 08/608,241
;
; EARLIER FILING DATE: 1996-02-28
;
; NUMBER OF SEQ ID NOS: 7
;
; SOFTWARE: PatentIn Ver. 2.1
;
; SEQ ID NO 3
;
; LENGTH: 4826
;
; TYPE: DNA
;
; ORGANISM: Rhodobacter sphaeroides
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (215)..(895)
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (993)..(2165)
;
; FEATURE:
;
; NAME/KEY: CDS
;
; LOCATION: (2236)..(4437)
;
; US-09-192-983-3

```

	Query Match	4.6%;	Score 33;	DB 4;	Length 4826;
	Best Local Similarity	46.7%;	Pred. No. 4.4;		
	Matches 105;	Conservative	0;	Mismatches 120;	Indels 0; Gaps.
Qy	288	CRAGGTGGAGCGCCACACGATCATTTGAAGAAACAGACACCAACCGTCCCGATGACACAGCC	347		
Db	3990	CATGCTGCAGGTGCCCCCGCCGCGAAAGTGCGCCGCTCCGTCCGTGCCCGATGCTCGGC	4049		
Qy	348	TTTGGGCGGTTTCGGCGGTGGCCACTGCACACGCGAACCGGTGCGGGTGAGGTGACCGT	407		
Db	4050	GGTTCGGGCGCAGGCCAAGGTGGCGGTGCAGACAAGATCGCTTCCTCGTCGAGAAATGA	4109		
Qy	408	CGATAGCAGCGGTTTGGGTAAACCCCATGTTGATGCAATCGTGCTGACCTGTTGAA	467		
Db	4110	CGACGACCTGCGTCAGCGCATGGGACTGCTGCTGGAGAAATGGGGGGGTGAGCGTGCTCGA	4169		
Qy	468	CCCGAATTCGGTATTTGGAGCGCGTTTGTGTTTATCGCGCGCGCTCGG	512		
Db	4170	CGCGCCCTCGGGCGAGAGCGCTCGCGCTGATCGAGAGATCGG	4214		

```

RESULT 12
US-09-214-278-4
; Sequence 4, Application US/09214278
; Patent No. 6291210
; GENERAL INFORMATION:
; APPLICANT: Sakano, Sei-ji
; APPLICANT: Itoh, Akira
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8576
; CURRENT APPLICATION NUMBER: US/09/214, 278
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 3955
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)..(3725)
; NAME/KEY: sig_peptide
; LOCATION: (12)..(89)
; NAME/KEY: mat_peptide
; LOCATION: (90)..(3725)
US-09-214-278-4

```

Query Match 4.6%; Score 32.8; DB 4; Length 3955;
Best Local Similarity 58.0%; Pred. No. 4.6;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGAGCGTCGATAGCAGCGGGTTGGGTAAGCCCAT 437
Db 3353 GAGGAGCGGTGCGCGGGGAGGAGCGCCAAACACCAAGTGGCGCCGCTCAACCCCAT 3412
QY 438 GTTGATGCAATCGTGTGACCTGTTGTAACCCGGAATGCG 477
Db 3413 CCGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3452

RESULT 13

US-08-882-046-3
; Sequence 3, Application US/08882046
; Patent No. 6136952

GENERAL INFORMATION:

APPLICANT: Li, Linheng
APPLICANT: Hood, Leroy
APPLICANT: Krantz, Ian D.
APPLICANT: Spinner, Nancy B.
TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
TITLE OF INVENTION: Nucleic Acids and Methods of Use
NUMBER OF SEQUENCES: 110
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,046
FILING DATE: 25-JUN-1997
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-UW 2637
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 4315 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 16..3460

US-08-882-046-3

Query Match 4.6%; Score 32.8; DB 3; Length 4315;
Best Local Similarity 58.0%; Pred. No. 4.8;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGAGCGTCGATAGCAGCGGGTTGGGTAAGCCCAT 437
Db 3090 GAGGAGCGGTGCGCGGGGAGGAGCGCCAAACACCAAGTGGCGCCGCTCAACCCCAT 3149

QY 438 GTTGATGCAATCGTGTGACCTGTTGTAACCCGGAATGCG 477
Db 3150 CCGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3189

RESULT 14

US-08-611-729A-7

; Sequence 7, Application US/08611729A
; Patent No. 6004924

GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David
APPLICANT: Henrique, Domingos M.P.
APPLICANT: Lewis, Julian H.
APPLICANT: Myat, Anna M.
APPLICANT: Fleming, Robert J.
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Mann, Robert S.
APPLICANT: Gray, Grace E.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/611,729A
FILING DATE: 06-MAR-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-037
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 332..4102

US-08-611-729A-7

Query Match 4.6%; Score 32.8; DB 3; Length 4483;
Best Local Similarity 58.0%; Pred. No. 4.9;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGAGCGTCGATAGCAGCGGGTTGGGTAAGCCCAT 437
Db 3163 GAGGAGCGGTGCGCGGGGAGGAGCGCCAAACACCAAGTGGCGCCGCTCAACCCCAT 3222

QY 438 GTTGATGCAATCGTGTGACCTGTTGTAACCCGGAATGCG 477
Db 3223 CCGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3262

RESULT 15

PCT-US93-06251-34/c

; Sequence 34, Application PC/TUS9306251
; GENERAL INFORMATION:

APPLICANT: Wickstrom, Eric and Rife, Jason P.
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
NUMBER OF SEQUENCES: 93
CORRESPONDENCE ADDRESS:

Search completed: April 27, 2003, 03:47:50
Job time : 10403.1 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 2249.77 Seconds
(without alignments)
11293.030 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1421_2293
Perfect score: 873
Sequence: 1 gtcgataagcagcggtttg.....ctgtgagctctggaaccgtag 873

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_srs.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_srs.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	873	100.0	2374	1	CGLYSEG	X96471 C.glutamici
2	873	100.0	2374	6	A93933	Sequence 2
C 3	873	100.0	333150	1	AP005277	AP005277 Coryneb
C 4	873	100.0	349980	6	AX127147	AX127147 Sequence
C 5	568	65.1	627	6	AX123538	AX123538 Sequence
C 6	404.2	46.3	993	6	AX067087	AX067087 Sequence
7	326	37.3	822	6	AX063771	AX063771 Sequence
8	326	37.3	822	6	AX244059	AX244059 Sequence
9	303	34.7	708	6	AX123539	AX123539 Sequence
10	193.6	22.2	1771	1	AB083133	AB083133 Coryneb
11	141.2	16.2	345783	1	AP003001	AP003001 Mesorhiz
C 12	136.6	15.6	8472	1	AE008076	AE008076 Agrobacte
C 13	136.6	15.6	13051	1	AE009111	AE009111 Agrobacte
C 14	133.6	15.3	15886	1	SC7H9	AL450223 Streptomy
15	115	13.2	306250	1	SME591788	AL591788 Sinorhizo
C 16	88.6	10.1	10871	1	AE011965	AE011965 Xanthomon
C 17	88.4	10.1	36583	1	SC5H1	AL049863 Streptomy
C 18	87.2	10.0	1525	1	PSEIGRA	M37389 Pseudomonas
C 19	86.2	9.9	12257	1	AE002063	AE002063 Deinococc
C 20	84.8	9.7	10336	1	AE005530	AE005530 Escherich
C 21	83.6	9.6	266658	1	AP002563	AP002563 Escherich
C 22	80.4	9.2	11024	1	AE000382	U28377 Escherichia
C 23	80.4	9.2	141744	1	ECU28377	AC020884 Mus muscu
C 24	79.4	9.1	110000	2	AC020884_0	AL627274 Salmonell
C 25	76.4	8.8	256050	1	AE008808	AE008808 Salmonell
C 26	74.8	8.6	22929	1	AE008808	AL646059 Ralstonia
27	73.4	8.4	190050	1	AL646059	AR012071 Sequence
28	71.6	8.2	5541	6	AR025196	AR025196 Sequence
29	71.6	8.2	5541	6	AR038653	AR038653 Sequence
* 30	71.6	8.2	5541	6	AR068500	AR068500 Sequence
31	71.6	8.2	5541	6	AX343066	AX343066 Sequence
32	71.6	8.2	5541	6	AX347518	AX347518 Sequence
33	71.6	8.2	5541	6	AX347541	AX347541 Sequence
34	71.6	8.2	5541	6	AX347564	AX347564 Sequence
35	71.6	8.2	5541	6	SC5F8	AL357613 Streptomy
36	65.6	7.5	33285	1	SC5F8	AX431662 Sequence
C 37	65.4	7.5	988	6	AX431662	AE004852 Pseudomon
38	63.4	7.3	10542	1	AE004852	U65741 Aeromonas s
39	61.6	7.1	4033	1	ASU65741	AL646067 Ralstonia
C 40	60	6.9	213050	1	AL646067	AE012410 Xanthomon
C 41	59.6	6.8	10747	1	AE012410	U34849 Mycobacteri
C 42	57	6.5	14844	1	MEU34849	AE007056 Mycobacte
C 43	57	6.5	14869	1	AE007056	I86263 Sequence 17
C 44	57	6.5	15239	6	I86263	Z74025 Mycobacteri
45	57	6.5	38500	1	MTCY39	

ALIGNMENTS

RESULT 1
CGLYSEG

LOCUS C.glutamici 2374 bp DNA linear BCT 18-MAR-2001
DEFINITION C.glutamici lysE and lysG genes.
X96471
ACCESSION X96471
VERSION X96471.1 GI:1729753
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine
exporter protein; Lysine governor.
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 2374)
AUTHORS Vrijic,M., Sahm,H. and Eggeling,L.

TITLE A new type of transporter with a new type of cellular function:
L-lysine export from *Corynebacterium glutamicum*
JOURNAL Mol. Microbiol. 22 (5), 815-826 (1996)
MEDLINE 97126810
PUBMED 8971704
REFERENCE 2 (bases 1 to 2374)
AUTHORS Vrljic, M.M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie
1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
FEATURES
source
1..2374
/organism="Corynebacterium glutamicum"
/strain="R127"
/db_xref="taxon:1718"
complement (82..954)
/gene="lysG"
complement (82..954)
/gene="lysG"
/note="Lysine governor"
/codon_start=1
/transl_table=11
/product="Lysine export regulator protein"
/protein_id="CAA65323.1"
/db_xref="GI:1729754"
/db_xref="SWISS-PROT:P94632"
/translation="MNPIQLDTLLSIIDEGSEFEGASLALISPSAVSRVKALEHHVG
RVLSRTQPAKATEARVEVQARKNVLQAETKAQLSGLRLAELPIITIAINADSLSTW
FPLVFNVAWSGATLTLEDEAHTLSLRGRDVLGAVTREANPVAGCEVVELGTMR
HLAIAFPSLRDAVMVDKLDWAMPVLRPGKVDLQDRDLDRVDPGVRERRSVLPS
ASGFGAIRGIGLWGLLPETQAPMLKAGEVILLDEIPIDTPMYWRWRSLRL
TDAVDAALEGURP"
1025..1726
/gene="lysE"
1025..1726
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="CAA65324.2"
/db_xref="GI:13397387"
/db_xref="SWISS-PROT:P94633"
/translation="MEIFITGILLGASLLISIPQNVLVKIGIKREGLIIVLVCLL
SDVFLFIAGTGLVDSNAPIVLDIMRWGGIAYLLWFAVMAKAMTINKVAPQILE
ETEPVDDTFLGSAVATDTRNRVEVSDKQVRVPMALVILWLNPNAYLDA
FVPIGGVAGQYGTGRTVFAAGAFASLIWFFLVGFAGAAALSRPLSSPKVWRWVNVV
AVMTALAIKLMNG"
complement (1723..2352)
/gene="orf3"
complement (1723..2352)
/gene="orf3"
/codon_start=1
/transl_table=11
/protein_id="CAA65325.1"
/db_xref="GI:1729756"
/db_xref="SPRMBL:P94634"
/translation="WVALRDIVASGKALVYISSYGPETLTAEEAEEMAEGLLIHQ
PYSIINRWEEFDDGNLLOSAANNGLVAFSPLAQGLITDKYLDIGIPBGSRASQ
GSLSEGNLVNIMVRKLNIDIAOERQGLAQMALAWVLRQEGYGDVTSALIGA
SSVEQLDNLSDLSLNNLEFSDALEATDEISHDAGINWAKATDSKTREN"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 873; DB 1; Length 2374;
Best Local Similarity 100.0%; Pred. No. 4.7e-185;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGATAAGCAGCGGTTTCGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGGTGG 60
Db 1421 GTCGATAAGCAGCGGTTTCGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGGTGG 1480
QY 61 AACCCGAATCGGTATTGGACGGTTTGTGTTATCGCGCGCGTCCGCGGCAATACGGC 120

Db 1481 AACCCGAATCGGTATTGGACGGTTTGTGTTATCGCGCGCGTCCGCGGCAATACGGC 1540
QY 121 GACACCGAGCGTGGATTTTCGCGCTGCGCGTTCGCGGCAAGCCCTGATCTGGTTCCCG 180
Db 1541 GACACCGAGCGTGGATTTTCGCGCTGCGCGTTCGCGGCAAGCCCTGATCTGGTTCCCG 1600
QY 181 CTGGTGGGTTTCGCGCAGCAGATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGGCGC 240
Db 1601 CTGGTGGGTTTCGCGCAGCAGATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGGCGC 1660
QY 241 TGGATCAACGTCGTCTGCGCAGTGTGATGACCGCATTCGCCATCAACTGATGTTGATG 300
Db 1661 TGGATCAACGTCGTCTGCGCAGTGTGATGACCGCATTCGCCATCAACTGATGTTGATG 1720
QY 301 GGTAGTCTTCGCGGTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATGCGCGGCTCG 360
Db 1721 GGTAGTCTTCGCGGTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATGCGCGGCTCG 1780
QY 361 TGGGAAATCTCATGATCGCTCCAACTCGCGGTGAGAAATCTCCAAAGTTGTTGATGAA 420
Db 1781 TGGGAAATCTCATGATCGCTCCAACTCGCGGTGAGAAATCTCCAAAGTTGTTGATGAA 1840
QY 421 TCAAGCTCTGTTGTCAGCTGCTCAACTGACGAGCAACCAATCAATGCACTGGTTCAGGTA 480
Db 1841 TCAAGCTCTGTTGTCAGCTGCTCAACTGACGAGCAACCAATCAATGCACTGGTTCAGGTA 1900
QY 481 TCGCGCCGCTACTCTCTTGTCTCGCGCAGCAGCCCATGCAAGCGCCCATCTGCGCAAGTGAC 540
Db 1901 TCGCGCCGCTACTCTCTTGTCTCGCGCAGCAGCCCATGCAAGCGCCCATCTGCGCAAGTGAC 1960
QY 541 TCGCGCGGTTCTGCGCGATGTCATGAGCTTTCGCGCAGCATATCAATATGTTGATGTTTC 600
Db 1961 TCGCGCGGTTCTGCGCGATGTCATGAGCTTTCGCGCAGCATATCAATATGTTGATGTTTC 2020
QY 601 AACATGCCCTCAGACAGGACTTACCTGCTGCGCGCGGGAACCCCTCTGGAATTCATCG 660
Db 2021 AACATGCCCTCAGACAGGACTTACCTGCTGCGCGCGGGAACCCCTCTGGAATTCATCG 2080
QY 661 AGATATTTGTCGTGACGAGGCGCTCGCGCAAGTGTGAGAAAGCAATGACGCGCAAGACCA 720
Db 2081 AGATATTTGTCGTGACGAGGCGCTCGCGCAAGTGTGAGAAAGCAATGACGCGCAAGACCA 2140
QY 721 TTGTTGGAGCTGACGCAACAGTTCTCACCGTCATCGCGCGGTTCTCCACCAACGA 780
Db 2141 TTGTTGGAGCTGACGCAACAGTTCTCACCGTCATCGCGCGGTTCTCCACCAACGA 2200
QY 781 TTAATCATGGAATAGCTTCGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCGCATGAAC 840
Db 2201 TTAATCATGGAATAGCTTCGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCGCATGAAC 2260
QY 841 TCAGCGCGCTCCGCTGTGAGCTCTGACCGGTAG 873
Db 2261 TCAGCGCGCTCCGCTGTGAGCTCTGACCGGTAG 2293
RESULT 2
A93933
LOCUS Sequence 2 from Patent WO9723597.
DEFINITION A93933
ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1 (bases 1 to 2374)
Vrljic, M. and Eggeling, L.
PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
ACTIVITY OF EXPORT CARRIERS
Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLIJC MARINA (DE)

FEATURES		Location/Qualifiers	
source		1..2374	
BASE COUNT		526 a 540 c 648 g 560 t	
ORIGIN		100.0%; Score 873; DB 6; Length 2374; Best Local Similarity 100.0%; Pred. No. 4.7e-185; Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Query Match		100.0%; Score 873; DB 6; Length 2374;	
Best Local Similarity		100.0%; Pred. No. 4.7e-185;	
Matches 873; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	GTCCATAGCAGCGGTTGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG	60
Db	1421	GTCCATAGCAGCGGTTGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG	1480
Qy	61	AACCCGAATCGGTATTGAGACCGGTTGTTGTTATCGCGCGCGTGGCGGCAATACGGC	120
Db	1481	AACCCGAATCGGTATTGAGACCGGTTGTTGTTATCGCGCGCGTGGCGGCAATACGGC	1540
Qy	121	GACACCGACCGGTGATTTTCGCGCTGGCGGCTTGGCGGCAAGCCTGATCTGTTCCCG	180
Db	1541	GACACCGACCGGTGATTTTCGCGCTGGCGGCTTGGCGGCAAGCCTGATCTGTTCCCG	1600
Qy	181	CTGTTGGGTTTCGCGGACGAGCATTTGTCAGCCGCGCTGTCAGCCCAAGGTGTGGCGC	240
Db	1601	CTGTTGGGTTTCGCGGACGAGCATTTGTCAGCCGCGCTGTCAGCCCAAGGTGTGGCGC	1660
Qy	241	TGGATCAACGTCGTCGTCGATGTTGTGATGACCGCATTTGGCCATCAACTGATCTGATG	300
Db	1661	TGGATCAACGTCGTCGTCGATGTTGTGATGACCGCATTTGGCCATCAACTGATCTGATG	1720
Qy	301	GGTTAGTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCGCAAAATGTTGATGCCCGCGTCG	360
Db	1721	GGTTAGTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCGCAAAATGTTGATGCCCGCGTCG	1780
Qy	361	TGGGAATCTCATGATCGCTCCAACTCGCGGTCGAGAACTCAAGTCTGAGTGA	420
Db	1781	TGGGAATCTCATGATCGCTCCAACTCGCGGTCGAGAACTCAAGTCTGAGTGA	1840
Qy	421	TCAGGCTGTTGTCAGTCTCAACTGACGAAGCACCATCAATGACGACTGGTCACCGTA	480
Db	1841	TCAGGCTGTTGTCAGTCTCAACTGACGAAGCACCATCAATGACGACTGGTCACCGTA	1900
Qy	481	TCGCGCGCTACTCTCTCTGTCGCGGACCCATGCAAGCGCATCTGCGCAAGTGAC	540
Db	1901	TCGCGCGCTACTCTCTCTGTCGCGGACCCATGCAAGCGCATCTGCGCAAGTGAC	1960
Qy	541	TGCGCGGTTCTGCGGATGTCATTGAGCTTGGGACCATATCAATATTGTTCACTTC	600
Db	1961	TGCGCGGTTCTGCGGATGTCATTGAGCTTGGGACCATATCAATATTGTTCACTTC	2020
Qy	601	AACATGCCCTCAGACAGGACTTACCCTGGCTGCGCGGGAACCTCTGGAATTCATCG	660
Db	2021	AACATGCCCTCAGACAGGACTTACCCTGGCTGCGCGGGAACCTCTGGAATTCATCG	2080
Qy	661	AGATATTGTCGTCAGCAGCCCTGCGCAAGTGTGAGAAAGCAATGAGCCCAAGACCA	720
Db	2081	AGATATTGTCGTCAGCAGCCCTGCGCAAGTGTGAGAAAGCAATGAGCCCAAGACCA	2140
Qy	721	TTGTTGACGAGTCTGCAACAAGTTCTACCGTCATCGCCCGGTTCCCTCCACCAACGA	780
Db	2141	TTGTTGACGAGTCTGCAACAAGTTCTACCGTCATCGCCCGGTTCCCTCCACCAACGA	2200
Qy	781	TTAATGATGAATAGCTTGGCTGATGATCAAGAGCGGGCAGCCCTCTCGGCATGAAC	840
Db	2201	TTAATGATGAATAGCTTGGCTGATGATCAAGAGCGGGCAGCCCTCTCGGCATGAAC	2260
Qy	841	TCAGCCGCTCTCGCTGTGAGCTCTGGACCGTAG	873
Db	2261	TCAGCCGCTCTCGCTGTGAGCTCTGGACCGTAG	2293
RESULT 3			

AP005277/c	AP005277	333150 bp	DNA	linear	BCT 08-AUG-2002
LOCUS	Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 4/10.				
DEFINITION	Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.				
ACCESSION	AP005277	BA000036			
VERSION	AP005277.1	GI:21323710			
KEYWORDS	Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.				
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.				
ORGANISM	Nakagawa,S.				
REFERENCE	Complete genomic sequence of Corynebacterium glutamicum ATCC 13032				
AUTHORS	Unpublished				
TITLE	2 (bases 1 to 333150)				
JOURNAL	Nakagawa,S.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.				
JOURNAL	Ltd., Tokyo Research Laboratories; 3-6-6, Aeahi-machi, Machida, Tokyo 194-8533, Japan (E-mail:snakagawa@xanagen.com, Tel:81-44-829-3031, Fax:81-44-813-1651)				
COMMENT	This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.				
FEATURES	Location/Qualifiers				
source	1..333150				
	/organism="Corynebacterium glutamicum ATCC 13032"				
	/strain="ATCC 13032"				
	/db_xref="taxon:196627"				
	/note="ATCC 13032"				
gene	complement(63..1292)				
	/gene="Cg10944"				
CDS	complement(63..1292)				
	/gene="Cg10944"				
	/codon_start=1				
	/transl_table=11				
	/product="Hypothetical membrane protein"				
	/protein_id="BAB98337.1"				
	/db_xref="GI:21323711"				
	/translation="MTNPTERNARLLIWGLQNIQDQIVAAKTVLPWLQAGAPG FLIALLVPIREAGSMLPQAATGWLRTQSRVSKVWIGNSQFVSALGIGVAALFLRG WALGITVIVLLAALSFRSMGSIASKQVKVIGKRGVLTGRATVGGVGLVAGL AIAIPGSHSPTRVLAIVIASPSWLFASIVFARIYAKPATPKNAPSANPWRRCI AALKDKAPRRFLVRSMVLTAALFIVALAAGSNGSIDSLGPELLASGLASWVG RISGTDHSHSKWAGALPGSIVLILVLSAPAPAQINTLPPLSFLLTLAHTA IRVARKTYMDMAEGDQRTYRVADANTLMGVLLIVGALSGLFIAIFGNEALLFLAAI GLGTISARGLKEYSAG"				
gene	1382..2917				
	/gene="Cg10945"				
CDS	1382..2917				
	/gene="Cg10945"				
	/codon_start=1				
	/transl_table=11				
	/product="Putative multicopper oxidases"				
	/protein_id="BAB98338.1"				
	/db_xref="GI:21323712"				
	/translation="MPELNRTFFKAGVLAATVGAQVLVACSSDDVRGVRGPRRL PIPADLGTREGSVHFALQAQGESQILPDVTTKTWGFNGHTLGLPTLVVKKGGDVHV DVNNLDMETTVHWHGKMLPAIADGPHSPITGPGQVTSPTVTVANDATLWHPHTHG LTGLHVRGLAGMIIVDEADTKLDLPREYGVDDIPLVMDHRFLEDGSLDEEDLPDL GLLGDTPANGITNAHPDATRRVRFRVLNGSNRFRYNLAESDTRTFQVIASDGLLD EPQRTTLAIGPGRWEIIVLEPEGEDVTLESVFEDNYGVDPDFVDFGMSDSFQL LTIITGSDDAQAQALPGVLVKSTEPDVIDATERFTIMTFSINDLQMDMRQDVVID HDQPEVMIIVTNDSDPHFVHDARFKLVKEGTVELEFNDKWDKVTGLPFGATIL AVEFGHYDPQWPYMYHCHMLYHEDQGMWGQFVIVEPGDEPAAVLGSGTSSIDSAG HAH"				
gene	3063..3983				
	/gene="Cg10946"				
CDS	3063..3983				
	/gene="Cg10946"				
	/note="PF00005:ABC transporter				
	TIGR00960:3a0501802: Type II (General) Secretory Pathway				
	(IISP) Family protein				

```
TIGR00968:3a0106s01: sulfate transport system permease
protein
TIGR01166:cbiO: cobalt transport protein ATP-binding
subunit
TIGR01184:ntrCD: nitrate transport ATP-binding subunits C
and D
TIGR01186:proV: glycine betaine/L-proline transport ATP
binding subunit
TIGR01187:potA: spermidine/putrescine ABC transporter
ATP-binding subunit
TIGR01188:draA: daunorubicin resistance ABC transporter
ATP-binding subunit
TIGR01189:cmhA: heme exporter protein CcmA
/codon_start=1
/transl_table=11
/product="ABC-type transporter, ATPase component"
/protein_id="BAB98339.1"
/db_xref="GI:21323713"
/translations="MTTTPAIDVTDLVRTYGDYTAVKGLNFHVQGEVFGLLGTNGAG
KSTLEVIIGLSAPSGTVIRISGLDVADRILRPELGMQSGLPQSQTVAETMDM
WHQCTYPAIKDVLADVDLHRENVKVGALSGEQRDLACALGDSIFLDEPT
GLDPSRHRHTWOLLDLKQGVWMLTTHYLEAEFLCDRIAIMNAGEIAVEGTIDE
LVAREKSIISFVLRGGOVELPVLGSAEIIIRDNNHVRIATTTLQHTLEILTAAETGI
ALGPPAKATLESVPMDIASLENTSLQTA"
4003..4836
/gene="Cg10947"
CDS
4003..4836
/gene="Cg10947"
/codon_start=1
/transl_table=11
/product="ABC-type transporter, permease components"
/protein_id="BAB98340.1"
/db_xref="GI:21323714"
/translations="MTTSTARGLOHATPERKTSFKTSLFKAELQFRNKTLIFM
ATVPVGPILLPLIGNGGAESANFDVFMVTLFVQFYTVLSMATRDERVGLKR
LFTGEARDIDICAIPEGALLTLEVTIPLLVGLGAPAPINLPIVPAVLIGLL
CSALAMTDSIFTRAEAAQMTSPMFLMANGLGSRFVFGSIVADILATYPPFAIS
DLVQIGWAGATFADSVGGVEAANFAGIFQDMLIPLGLILAAVTAANRYNRWDSY
R"
4879..6138
/gene="Cg10948"
CDS
4879..6138
/gene="Cg10948"
/EC_number="2.7.3.-"
/codon_start=1
/transl_table=11
/product="Two-component system, sensory transduction
histidine kinases"
/protein_id="BAB98341.1"
/db_xref="GI:21323715"
/translations="MNKDFWTAGWTARWFSRGVSLASPVTAFLNSWRRLPNLAKYTL
YTRVSLQAIPLVLLSAYFLGIVANAGTNPSFVMLGFSVILLIVTLVVEYQPSLNS
HPRSVPQPFPTGLNLVNLGVSVVLOIPGLANSNDTRATLIFTLCVFLLSIAYI
PMNRYWMLAMSALVLMWTSTTVDVLSALMWVIPLMAGTVLSVTVTVDMKEVERS
RELASLRVTEERLRFAQLHDLTGLQHLAAMSVKSHALAKRGGDRLENELRELQK
LRTSMSEMDYVSGRYVTVNLATEIEGAKSLADAHILHSVIGTTSQVSPAPHELCAW
LVREATTNLRHSDATDATLTLSSTEVRMDNNGVKNKIDIGRLSGLSALRSRAESAGMTL
IVSRDDQFVSRLINAPANTPAEKA"
6139..6750
/gene="Cg10949"
CDS
6139..6750
/gene="Cg10949"
/notes="PF000072:Response regulator receiver domain
PF00196:Bacterial regulatory proteins, luxR family"
/codon_start=1
/transl_table=11
/product="Two-component system, response regulators
consisting of a CheY-like receiver domain and a HTH
DNA-binding domain"
/protein_id="BAB98342.1"
/db_xref="GI:21323716"
/translations="MISISIAADDEALIASLATLLSLEPDLVDVRPTAGSGEELIETWA
DPSNRDVCVLDLQGLDGDIDTRMLMETTPDLAVLITVSHARPQLKRALAAGVLG
```

```
FLPKTSTADEFAFAIRTVHAGRYIDPELAAMTISAGESPLTNEEEVLELAGQGLSA
BEIIVAAHLAPGTTNRNLYSQMTKVGAGNRFEATRAHELGLW"
complement (6781..6987)
/gene="Cg10950"
complement (6781..6987)
/gene="Cg10950"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="BAB98343.1"
/db_xref="GI:21323717"
/translations="MKHYWSMYFGAGPLNTPHTKEVWFMDIVSVIWNHFWFSTNNWVN
MTGLPRETFVPAFGSTVLAFLSF"
7048..8205
/gene="Cg10951"
CDS
7048..8205
/gene="Cg10951"
/codon_start=1
/transl_table=11
/product="crna-processing ribonuclease BN"
/protein_id="BAB98344.1"
/db_xref="GI:21323718"
/translations="MTAWLCSPSGVQDEYDMNGVQVOPQSHLDATLIAADPHGNPEN
SGDKERLNFQGWKVALNRTVDPDPGLDLAALLTFPSILSTAPAVLIGVSVITIF
LASSTELNLNVRDEVNQYVEDQSHVNVGVDSIAGSAAAGQGVAVGVITALTWTS
AYVRAFRCANAVRSEGRSLIKRWMLLFNLALLGLIILVSVVNLNLTVMGIF
APIAPLPLTNLVSFLTDREPMIWIWRFVPIVGLIMFVATLYYAPNAPRPFKRWL
SLGSLAIVGILLAGVGLNFYFTLFAAFSSYGAVGSLIAVFIALVFNICLIIGLKID
VEISRKOLQAMPADYSLVPRSRSEKIVAKMKQORQLMDQAAAAIREESN"
8672..10114
/gene="Cg10952"
CDS
8672..10114
/gene="Cg10952"
/notes="PF000005:ABC transporter
PF00664:ABC transporter transmembrane region."
/codon_start=1
/transl_table=11
/product="ABC-type multidrug/protein/lipid transport
system, ATPase component"
/protein_id="BAB98345.1"
/db_xref="GI:21323719"
/translations="MMRTTHTASSILRRMIRROKQKAVGAGFFLGMWLSBALVPFIAI
GLIVDHAIVTDLRELVLGVLFVFLVSVFSGFSGRNLNRNVPFESHARVERVAD
HALKMLDPNLVPGVMSRSTADSDSTRIFQIGTGTGSAATGFLCAATYLLISDLV
GLLVLPVLIISGVVALASGKISKRSTQQLKLAESGAQSDIMMLGLRVIKAIIGERW
AVKTEKASQASARAADVTAVASGKAGIGELSIANVLAVLLAGWRTVTGELGPGQ
LIAIVGVAVLSEPIRLLSNSINASIAHGAARVANFLNDESQAQYESSSETINDGE
FLIVVPASTLPHGDNILATPHAADIPEGLTNRNISMNHDNVPIDPOVIRASGLTDI
LEVDSLDAPRDTGSNLSGGQQRVALARALHADAEVLVLMDDPTSAVDSVTEVSTIAQG
IKQLRAGKTTIVSSSPAFYNLADRIVSHV"
10107..11843
/gene="Cg10953"
CDS
10107..11843
/gene="Cg10953"
/notes="PF000005:ABC transporter"
Query Match 100.0%; Score 873; DB 1; Length 333150;
Best Local Similarity 100.0%; Pred. No. 4e-185;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCAATCGTGACCTGGTTG 60
|||||
Db_328598 GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCAATCGTGACCTGGTTG 328599
QY 61 AACCCGAATCCGTATTGGACGGCTTTGTGTTTATCGGCGCGCTCGCGCAATACCGC 120
|||||
Db_328538 AACCCGAATCCGTATTGGACGGCTTTGTGTTTATCGGCGCGCTCGCGCAATACCGC 328479
QY 121 GACACCGGACGGTGGATTTTCGCCGCTTGGCGCGCTTTCGCCGCAAGCCGTATCTGGTCCCG 180
|||||
Db_328478 GACACCGGACGGTGGATTTTCGCCGCTTGGCGCGCTTTCGCCGCAAGCCGTATCTGGTCCCG 328419
QY 181 CTGGTGGGTTTCGCCGACGACGACATCTGTACAGCCCGCTGTCCAGCCCCCAAGGTGTGCGCG 240
```



```

Db 328418  CTGGTGGGTTTTCGGCGCAGCAGCATTGTACGCGCGCTGTCCAGCCCAAGGTGTGGCGC 328359
QY 241  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATTGGCCATCAAACTCATGTTGATG 300
Db 328358  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATTGGCCATCAAACTCATGTTGATG 328299
QY 301  GGTAGTTTTTCGGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGCTCG 360
Db 328298  GGTAGTTTTTCGGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGCTCG 328239
QY 361  TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAGTCAAGAACTCAAGTTGTTGAGTGAA 420
Db 328238  TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAGTCAAGAACTCAAGTTGTTGAGTGAA 328179
QY 421  TCAAGGCTGTTGTCAGCTGTCTCAACTGACGAAGCACCACCAATCAATGCACTGTCACCGTA 480
Db 328178  TCAAGGCTGTTGTCAGCTGTCTCAACTGACGAAGCACCACCAATCAATGCACTGTCACCGTA 328119
QY 481  TCCGCGCGCTACTCTCTCTGTCGCGCAGCACCACCAATGCAAGCGCCATCTGCGCAAGTGAC 540
Db 328118  TCCGCGCGCTACTCTCTCTGTCGCGCAGCACCACCAATGCAAGCGCCATCTGCGCAAGTGAC 328059
QY 541  TGCCCGGCTTCTGGCGATGTCATTGAGCTTGGCGGACCATATCAATATTGTTCAAGTTC 600
Db 328058  TGCCCGGCTTCTGGCGATGTCATTGAGCTTGGCGGACCATATCAATATTGTTCAAGTTC 327999
QY 601  AACATGCGCTCAGACAGGAGCTTACCCCTGGCTGGCGGGGAACTCTGGAATTCATCG 660
Db 327998  AACATGCGCTCAGACAGGAGCTTACCCCTGGCTGGCGGGGAACTCTGGAATTCATCG 327939
QY 661  AGATATTGTCGCTGAGCAGCCCTTCGCGAAGTGTGAGAAAGCAATGACGCCAAGACCA 720
Db 327938  AGATATTGTCGCTGAGCAGCCCTTCGCGAAGTGTGAGAAAGCAATGACGCCAAGACCA 327879
QY 721  TTGTTGGCAGCTGACTGCAACAGTTCTACCGTCATCGCCCGGTCTTCCACCAAGCA 780
Db 327878  TTGTTGGCAGCTGACTGCAACAGTTCTACCGTCATCGCCCGGTCTTCCACCAAGCA 327819
QY 781  TTAATGATGGAATAGCTTGGCTGATGATGATCAGAAGCGGCGAGCCCTCTCGCCATGAC 840
Db 327818  TTAATGATGGAATAGCTTGGCTGATGATGATCAGAAGCGGCGAGCCCTCTCGCCATGAC 327759
QY 841  TCAGCGCGCTCCGCTGTGAGCTCTGGACCGTAG 873
Db 327758  TCAGCGCGCTCCGCTGTGAGCTCTGGACCGTAG 327726

```

```

RESULT 4
AX127147/c
LOCUS 349980 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 7063 from Patent EP1108790.
ACCESSION AX127147 AX114121
VERSION AX127147.1 GI:14041135.
KEYWORDS
SOURCE
ORGANISM

```

```

Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 349980)
Nakagawa, S., Mizoguchi, H., Ando, S., Hayaishi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 7063 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)

```

```

FEATURES
source
1. 349980
/organism="Corynebacterium glutamicum"
/db xref="taxon:1718"
/note="Seq 1 to long (3,309,400) split in 11, seq 7063
1.200.001 1.549.980"
BASE COUNT 80727 a 91049 c 97618 g 80586 t

```

ORIGIN

```

Query Match 100.0%; Score 873; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 4e-185;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTG 60
Db 128548  GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTG 128489

QY 61  AACCCGAATGCGTATTTGGACGCGTTTGTATTCGGCGGCGTTCGCGCGCAATACGCG 120
Db 128488  AACCCGAATGCGTATTTGGACGCGTTTGTATTCGGCGGCGTTCGCGCGCAATACGCG 128429

QY 121  GACACCGGACGGTGGATTTTCGCCGCTGGCGGCTTCGCGGCAAGCTGATCTGGTTCCCG 180
Db 128428  GACACCGGACGGTGGATTTTCGCCGCTGGCGGCTTCGCGGCAAGCTGATCTGGTTCCCG 128369

QY 181  CTGGTGGGTTTTCGGCGCAGCAGCATTTGTCAAGCCCGCTGTCCAGCCCCAAGGTGTGGCG 240
Db 128368  CTGGTGGGTTTTCGGCGCAGCAGCATTTGTCAAGCCCGCTGTCCAGCCCCAAGGTGTGGCG 128309

QY 241  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 300
Db 128308  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 128249

QY 301  GGTAGTTTTTCGGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGTTCG 360
Db 128248  GGTAGTTTTTCGGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGTTCG 128189

QY 361  TGGGAATCTCATGATCGCTCCAACTCGCGGTTCAGAAAACTCCAAAGTTGTTGAGTGAA 420
Db 128188  TGGGAATCTCATGATCGCTCCAACTCGCGGTTCAGAAAACTCCAAAGTTGTTGAGTGAA 128129

QY 421  TCAGCTGTTGTCAGCTGTCTCAACTGACGAAGCAACCAATCAATGCACTGTGTCAGGTA 480
Db 128128  TCAGCTGTTGTCAGCTGTCTCAACTGACGAAGCAACCAATCAATGCACTGTGTCAGGTA 128069

QY 481  TCCGCGCGCTACTCTCTGCTCGCGCAGCACCCTGCAAGCGCCATCTGCGCAAGTGAC 540
Db 128068  TCCGCGCGCTACTCTCTGCTCGCGCAGCACCCTGCAAGCGCCATCTGCGCAAGTGAC 128009

QY 541  TGCCCGGCTTCTGGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTGTTGAGTTC 600
Db 128008  TGCCCGGCTTCTGGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTGTTGAGTTC 127949

QY 601  AACATGCCCTCAGACAGGCACTTACCTGGCTGGCGGCGGAAACCTCTGGAATTCATCG 660
Db 127948  AACATGCCCTCAGACAGGCACTTACCTGGCTGGCGGCGGAAACCTCTGGAATTCATCG 127889

QY 661  AGATATTTGTCGTCAGCAGGCGCTTCGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 720
Db 127888  AGATATTTGTCGTCAGCAGGCGCTTCGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 127829

QY 721  TTGTTGGCAGCTGACTGCAACAAAGTTCTCACCGTCATCGCCCGGTTCCTCCACCAAGCA 780
Db 127828  TTGTTGGCAGCTGACTGCAACAAAGTTCTCACCGTCATCGCCCGGTTCCTCCACCAAGCA 127769

QY 781  TTAATGATGGAATAGCTTGGCTGATGATGATCAGAAGCGGCGAGCCCTCTCGCCATGAC 840
Db 127768  TTAATGATGGAATAGCTTGGCTGATGATGATCAGAAGCGGCGAGCCCTCTCGCCATGAC 127709

QY 841  TCAGCGCGCTCCGCTGTGAGCTCTGACCGTAG 873
Db 127708  TCAGCGCGCTCCGCTGTGAGCTCTGACCGTAG 127676

```

```

RESULT 5
AX123538/c
LOCUS 627 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3454 from Patent EP1108790.
ACCESSION AX123538
VERSION AX123538.1 GI:14041026

```

KEYWORDS
SOURCE
ORGANISM

Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 627)
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K., Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3454 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source 1. .627
Location/Qualifiers
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 139 a 159 c 185 g 144 t
ORIGIN

Query Match 65.1%; Score 568; DB 6; Length 627;
Best Local Similarity 100.0%; Pred. No. 8.9e-117;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 306 GTTTTCGGGGTTTGAATCGGTGGCTTCGCCCAATGTTGATCGCGGCTCGTGGGA 365
DB 627 GTTTTCGGGGTTTGAATCGGTGGCTTCGCCCAATGTTGATCGCGGCTCGTGGGA 568
QY 366 AATCTATCGATCGCTCCCACTCGCGTTCAGAAAATCCAAAGTTGTTGATGAATCAAG 425
DB 567 AATCTATCGATCGCTCCCACTCGCGTTCAGAAAATCCAAAGTTGTTGATGAATCAAG 508
QY 426 GCTGTTTCAGCTGCTCAACTGACGAGACCAATCAATGCACTGCTGATCCGC 485
DB 507 GCTGTTTCAGCTGCTCAACTGACGAGACCAATCAATGCACTGCTGATCCGC 448
QY 486 GCCGTACTCTCTTCTGCGGCGAGCACCATGCAAGCGCCATCTGCGCAAGTGAATGCCC 545
DB 447 GCCGTACTCTCTTCTGCGGCGAGCACCATGCAAGCGCCATCTGCGCAAGTGAATGCCC 388
QY 546 GCCTTCTGCGGCGAGTCAATGAGCTTTCGCGACCATATCAATATTTGTTCAAGTTCAACAT 605
DB 387 GCCTTCTGCGGCGAGTCAATGAGCTTTCGCGACCATATCAATATTTGTTCAAGTTCAACAT 328
QY 606 GCCCTCAGACAGGACTTACCTGCTGGCGGGAGACCTCTGGAATTCATCGAGATA 665
DB 327 GCCCTCAGACAGGACTTACCTGCTGGCGGGAGACCTCTGGAATTCATCGAGATA 268
QY 666 TTGCTCGTGAGAGGCGCTCGCGAAGTGGTGAGAAAGCAATGACGCCAAGCATTGTT 725
DB 267 TTGCTCGTGAGAGGCGCTCGCGAAGTGGTGAGAAAGCAATGACGCCAAGCATTGTT 208
QY 726 GGCAGCTGACTGCAACAGTTTCTACCGTCAATCGCCGGTTTCTCCACCCCAAGATTAAAT 785
DB 207 GGCAGCTGACTGCAACAGTTTCTACCGTCAATCGCCGGTTTCTCCACCCCAAGATTAAAT 148
QY 786 GATGAATAGCTTGGCTGATCAATCAGAGCGGCGAGCCCTCTCCGCGCATGAATCAGC 845
DB 147 GATGAATAGCTTGGCTGATCAATCAGAGCGGCGAGCCCTCTCCGCGCATGAATCAGC 88
QY 846 CGCCTCCGCTGTGAGCTCTGGACCGGTAG 873
DB 87 CGCCTCCGCTGTGAGCTCTGGACCGGTAG 60

RESULT 6
AX067087/c
LOCUS AX067087 993 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 669 from Patent WO010805.
ACCESSION AX067087
VERSION AX067087.1 GI:12544795
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum.

REFERENCE 1 (bases 1 to 993)
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE Corynebacterium glutamicum genes encoding proteins involved in membrane synthesis and membrane transport
JOURNAL Patent: WO 010805-A 669 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
source 1. .993
Location/Qualifiers
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
CDS
101. .973
/note="RXN03164"
/codon_start=1
/transl_table=11
/protein_id="CAC26771.1"
/db_xref="GI:12544796"

/translation="MIYRRVNSGLKLPALISGLWHNFGDDKPLSTQSIHRAFDGR
VTHFDLANNYPGPAAGSNETNFRILREDLKSRLDELIISKAGDMWMPYGFSGSRK
YLVSLDQSLTRLGLDYVDIFYHRRPDPTLEETMYALRDIYVAGKALYVSIYSGP
ELTAAEFMAEBCPLLIHOPYSYIINRWVEEPDGDGDNLIQSAANNGLGVIAFSP
AQGLTDKVLGIPGEGSRASQKSLSEGMLNVNINIDMVKLNDIAOERQSLAQWALA
WVLRQREYVGLP"
BASE COUNT 222 a 247 c 277 g 247 t
ORIGIN
Query Match 46.3%; Score 404.2; DB 6; Length 993;
Best Local Similarity 98.8%; Pred. No. 4e-80;
Matches 418; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
QY 453 AGCACCAATCAATGCACTGCTCAOGGTA--TCGGGCGGTACTCTCTTGTGTCGGCAGC 510
DB 993 AGCACCAATCAATGCACTGCTCAOGGTAATCGGGCGCGGTACTCTCTTGTGTCGGCAGC 934
QY 511 ACCCATGCAAGCGCCATCTGCGCAAGTGAATGCGCGGTTCTTGGCGGATGATGAGC 570
DB 933 ACCCATGCAAGCGCCATCTGCGCAAGTGAATGCGCGGTTCTTGGCGGATGATGAGC 874
QY 571 TTGCGGACCATATCAATATTTGTTTCAACATGCGCCCTCAGACAGGACTTACCCCTGG 630
DB 873 TTGCGGACCATATCAATATTTGTTTCAACATGCGCCCTCAGACAGGACTTACCCCTGG 814
QY 631 CTGGCGGGGAACCCCTCTGGAATTCATCGAGATATTTGTCCGTGAGCAGGCCCTGCGCA 690
DB 813 CTGGCGGGGAACCCCTCTGGAATTCATCGAGATATTTGTCCGTGAGCAGGCCCTGCGCA 754
QY 691 AGTGTGTGAGAAAGCAATGACGCCAAGCATTGTTGGCAGCTGACTGCAACAGTTCTCA 750
DB 753 AGTGTGTGAGAAAGCAATGACGCCAAGCATTGTTGGCAGCTGACTGCAACAGTTCTCA 694
QY 751 CCCTCATCGCCCGGTTCTCCACCCCAACGATTAATGATGAATAGCTTGGCTGATGAATC 810
DB 693 CCCTCATCGCCCGGTTCTCCACCCCAACGATTAATGATGAATAGCTTGGCTGATGAATC 634
QY 811 AGAAGCGGGCAGCCCTCTCCGCGCATGAATCAGCCGCTCCGCTGAGCTCTGAGCCG 870
DB 633 AGAAGCGGGCAGCCCTCTCCGCGCATGAATCAGCCGCTCCGCTGAGCTCTGAGCCG 574
QY 871 TAG 873
DB 573 TAG 571

RESULT 7
AX063771
LOCUS AX063771 822 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 53 from Patent WO010843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
SOURCE Corynebacterium glutamicum.

ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 822)
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
corynebacterium glutamicum genes encoding metabolic pathway
proteins
JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;
BAS F AKTIENGESSELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..822
CDS /organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..802
/note="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC25127.1"
/db_xref="GI:12541484"
SDVFLIAGTGLVLLNAAPIVLDIMRWGSIYLLWFAVMAAKDAMTKVAPQIIIE
ETEPTVDDPLGSSAVATDTRNRVRVESVDKQVWVKPMLMAVLTLNPNAYLDA
FVFIGGVAQYGDTRWIFAFAGAAASLIWFLVFGAAALSRLSPKVMRWNVVV
AVMTALAIAKMLMG"

BASE COUNT 167 a 192 c 246 g 217 t

ORIGIN

Query Match 37.3%; Score 326; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.3e-62;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 60
DB 497 GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 556

QY 61 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTGCGCGCAATACGGC 120
DB 557 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTGCGCGCAATACGGC 616

QY 121 GACACCGACGCGTATTTCGCGCTGCGCGCTGCGCGCAATGCTGTTCCCG 180
DB 617 GACACCGACGCGTATTTCGCGCTGCGCGCTGCGCGCAATGCTGTTCCCG 676

QY 181 CTGGTGGGTTTCGCGCAGCAGCATTGTCCAGCCGCTGTCAGCCCAAGCTGATCTGGTCCCG 240
DB 677 CTGGTGGGTTTCGCGCAGCAGCATTGTCCAGCCGCTGTCAGCCCAAGCTGATCTGGTCCCG 736

QY 241 TGGATCAACGTCGTGGCGAGTGTGATGACCGCATTCGCCATCAAACTGATGTTGATG 300
DB 737 TGGATCAACGTCGTGGCGAGTGTGATGACCGCATTCGCCATCAAACTGATGTTGATG 796

QY 301 GGTAGTTTTCGCGGTTTGGATC 326
DB 797 GGTAGTTTTCGCGGTTTGGATC 822

RESULT 8
AX244059
LOCUS AX244059 822 bp DNA linear PAT 28-SEP-2001
DEFINITION Sequence 51 from Patent WO0166573.
ACCESSION AX244059
VERSION AX244059.1 GI:15859123
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 822)
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O., Haberhauer, G.,
Kim, J.W., Lee, H.S. and Hwang, B.J.
Corynebacterium glutamicum genes encoding metabolic pathway

proteins
JOURNAL Patent: WO 0166573-A 51 13-SEP-2001;
BAS F AKTIENGESSELLSCHAFT (DE)
FEATURES Location/Qualifiers
source 1..822
CDS /organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..802
/note="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC88508.1"
/db_xref="GI:15859124"
/translation="MEIFITGLLGASILLISGPONVLVIKQIKREGILAVLLVCLII
SDVFLIAGTGLVLLNAAPIVLDIMRWGSIYLLWFAVMAAKDAMTKVAPQIIIE
ETEPTVDDPLGSSAVATDTRNRVRVESVDKQVWVKPMLMAVLTLNPNAYLDA
FVFIGGVAQYGDTRWIFAFAGAAASLIWFLVFGAAALSRLSPKVMRWNVVV
AVMTALAIAKMLMG"

BASE COUNT 167 a 192 c 246 g 217 t

ORIGIN

Query Match 37.3%; Score 326; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.3e-62;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 60
DB 497 GTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 556

QY 61 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTGCGCGCAATACGGC 120
DB 557 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTGCGCGCAATACGGC 616

QY 121 GACACCGACGCGTATTTCGCGCTGCGCGCTGCGCGCAATGCTGTTCCCG 180
DB 617 GACACCGACGCGTATTTCGCGCTGCGCGCTGCGCGCAATGCTGTTCCCG 676

QY 181 CTGGTGGGTTTCGCGCAGCAGCATTGTCCAGCCGCTGTCAGCCCAAGCTGATCTGGTCCCG 240
DB 677 CTGGTGGGTTTCGCGCAGCAGCATTGTCCAGCCGCTGTCAGCCCAAGCTGATCTGGTCCCG 736

QY 241 TGGATCAACGTCGTGGCGAGTGTGATGACCGCATTCGCCATCAAACTGATGTTGATG 300
DB 737 TGGATCAACGTCGTGGCGAGTGTGATGACCGCATTCGCCATCAAACTGATGTTGATG 796

QY 301 GGTAGTTTTCGCGGTTTGGATC 326
DB 797 GGTAGTTTTCGCGGTTTGGATC 822

RESULT 9
AX123539
LOCUS AX123539 708 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3455 from Patent EP1108790.
ACCESSION AX123539
VERSION AX123539.1 GI:14041027
KEYWORDS
SOURCE
ORGANISM Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 708)
Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES Location/Qualifiers
source 1..708
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"

BASE COUNT 134 a 173 c 221 g 180 t

ORIGIN

Query Match 34.7%; Score 303; DB 6; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.8e-57;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAGCAGCGGTTGGTAAGCCCATGTTGATGCAATCGTGCACCTGTTG 60
DB 406 GTCGATAGCAGCGGTTGGTAAGCCCATGTTGATGCAATCGTGCACCTGTTG 465

QY 61 AACCCGAATCGTATTGGACCGCTTTGTGTTATCGCGCGCTCGCGGCAATACGGC 120
DB 466 AACCCGAATCGTATTGGACCGCTTTGTGTTATCGCGCGCTCGCGGCAATACGGC 525

QY 121 GACACCGACCGGTGATTTTCGCCGCTGCGCGCTTCGCGGCAACCTGATCTCCCG 180
DB 526 GACACCGACCGGTGATTTTCGCCGCTGCGCGCTTCGCGGCAACCTGATCTCCCG 585

QY 181 CTGCTGGGTTTCGGCGAGCAGCATTTGTCACGCGCTGTCACGCCCAAGTGTGGCGC 240
DB 586 CTGCTGGGTTTCGGCGAGCAGCATTTGTCACGCGCTGTCACGCCCAAGTGTGGCGC 645

QY 241 TGGATCAACGTCGTCGTCGCGATTTGATGACCGCATTTGGCCATCAAACTGATCTGATG 300
DB 646 TGGATCAACGTCGTCGTCGCGATTTGATGACCGCATTTGGCCATCAAACTGATCTGATG 705

QY 301 GGT 303
DB 706 GGT 708

RESULT 10

AB083133
LOCUS
DEFINITION
Corynebacterium efficiens lysG, lysE genes for Lysine export transcriptional regulatory protein, Lysine exporter protein, complete cds.
1771 bp DNA linear BCT 06-APR-2002

ACCESSION
AB083133
VERSION
AB083133.1 GI:20065738
KEYWORDS
Corynebacterium efficiens DNA.
SOURCE
Corynebacterium efficiens
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE

1 Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
lysG, lysE of Corynebacterium efficiens
2 Published Only in Database (2002)
Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
Submitted (04-APR-2002) Hiroshi Itaya, AJINOMOTO CO., INC, Fermentation & Biotechnology Laboratories; 1-1, Suzuki-cho, Kawasaki-Ku, Kawasaki, Kanagawa 210-8681, Japan
(E-mail:hiroshi.itaya@ajinomoto.com, Tel:81-44-244-7123 (ex. 4146), Fax:81-44-222-0129)

FEATURES

source
1..1771
/organism="Corynebacterium efficiens"
/db_xref="taxon:152794"
complement(47..919)
/gene="lysG"
complement(47..919)
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine export transcriptional regulatory protein"
/protein_id="BAB88826.1"
/db_xref="GI:20065739"
/translation="MNPILDTLTIIDEGSFENASIALSISPSAVSQRKALEKSVG
RVLVSRTQPAVATGAEVLVQARKMALQAEITQEAERLDIPLTVAINADSLSTM
FPVFAEVHGWAVTLRLVEDEAHTLSLLRRGSLGAVTREADPVAGCEVRLGWVR

HLPVATPELRARYTVDGPDVVRMPVLFEPGNDVLQDELDLEGRVDGAVARRVSVVPS
AEGFGEAVRLGGLGMLPEAQAPMLAAGDVQLDEKVVDTPLVQWRLESRLRL
TDAVVDAARAGLRT"
981..1676
/gene="lysE"
981..1676
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="BAB88827.1"
/db_xref="GI:20065740"
/translation="MRDMEIFVTLGLIGASLLAIGPONVLVIKQIKREGITAVIIV
CLLSVVLFTGLTGLVGLISDTAPILDLRCWGIAYLLFAPVMAARDALRARTVTF
VHSEFPAASASGGVITTKORPRITRTGVRVWRPMLMAVITLWLNPNYLDLAFV
FEGGVAQYGETGRWIFPAAGAPASLVWFFLVGVGAALSRPLSPRVWRWNIIGVAV
VLTLGLAVKLILMG"

BASE COUNT 294 a 603 c 545 g 329 t
ORIGIN

Query Match 22.2%; Score 193.6; DB 1; Length 1771;
Best Local Similarity 78.4%; Pred. No. 5.4e-33;
Matches 232; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 10 CAGCGGTTTGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGACCCGAAT 69
DB 1380 CGCAGGTCGTGGTTCAGGCCCATGCTATGGCCATTTGCTGACCTGGCTCAATCCCAAT 1439

QY 70 GCGTATTGGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACCGCCACACCGGA 129
DB 1440 GCTACTAGATGCTTCTGTTCTTCTGATCGGTGTGCGAGCCAGTACGGGAGACCGGT 1499

QY 130 CGGTGATTTTCGCCGCTCGCGCTTCGCGGCAAGCTGATCTGCTCCGCTGGTGGGT 189
DB 1500 CGGTGATTTTCGCCGCTCGCGCTTCGCGGCAAGCTGATCTGCTCCGCTGGTGGGT 1559

QY 190 TTCGGCGCAGCAGCATTTGACGCGCTTCGCGGCAAGCTGATCTGCTCCGCTGGTGGGT 249
DB 1560 TACGGCGCGCGCATTTGACGCGCTTCGCGGCAAGCTGATCTGCTCCGCTGGTGGGT 1619

QY 250 GTGCTGCTGGCAGTTCGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGTTA 305
DB 1620 ATAGTGTGGCGCTGTGCTGCTACCGGATTTGGCGTGAAGTATCTGATGGTTA 1675

RESULT 11

AP003001
LOCUS
DEFINITION
Mesorhizobium loti DNA, complete genome, section 8/21.
ACCESSION
AP003001 BA000012
VERSION
AP003001.2 GI:14023393
KEYWORDS
Mesorhizobium loti (strain:MAFF303099) DNA.
SOURCE
Mesorhizobium loti
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Phyllobacteriaceae; Mesorhizobium.

REFERENCE

1 (sites)
AUTHORS
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Iidesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuoka, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
MEDLINE
21082930
REFERENCE
2 (bases 1 to 345783)
AUTHORS
Kaneko, T.
Direct Submission
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp)

URL: <http://www.kazusa.or.jp/rhizobase/>,
 Tel: 81-438-52-3355 (ex.2338), Fax: 81-438-52-3934)
 On May 11, 2001 this sequence version replaced gi:11994976.

Location/Qualifiers

1. 345783
 /organism="Mesorhizobium loti"
 /strain="MAFP303099"
 /db_xref="taxon:381"
 complement(6. .884)
 /gene="mll3011"
 complement(6. .884)
 /gene="mll3011"
 /codon_start=1
 /transl_table=11
 /product="transcription regulator"
 /protein_id="BAB50000.1"
 /db_xref="GI:14023394"
 /translations="MDRLAMSTFIRVIETGSPSAAARHLRVGQSAVKVAQAOLEERL
 GVLLSRTEGLSPTDAGLSFYSGAKAVDEADEAANGTAAALTGLLRTIGTSTTAA
 GRLLVPRLKAFDAIDAPRLDVEAMDDRDVLEEGVDVMRLGESFPAPAHIGRKAIA
 CRTLVAAAPSYLAREGEPIADPLESRHEALIFVRAGTSIEWTRRGAEEVSLAQOGRLL
 RVTSGAIDREAVLSGIGFTVGSSEMFDAELRSGAVRPILTEWTLPPVELWALFPAGRK
 ASARARTVEFFLAGSL"
 994. 1740
 /gene="mlr3013"
 994. 1740
 /gene="mlr3013"
 /codon_start=1
 /transl_table=11
 /product="short chain dehydrogenase or 3-ketoacyl-CoA
 reductase"
 /protein_id="BAB50001.1"
 /db_xref="GI:14023395"
 /translations="MGKLNKKVAITGAARGIGAGIAKRLAAEGAAVVMYATSRRAA
 ERTVKIVAAAGKATAVQGVDAEAEVKMFEEAETAFGKPSILINNAATASPTGEE
 ASADYRRLPDTNVLGTILTTKAAHPFETGGSIVNIGTISWNPVNTSLTYSASKA
 AIDTLTSLRSELGARNIRNIVAPGTTHTEMTGEMVIGTIDFGNNLIAGVPLGQRFQK
 DDIAPIVAFIADSEAAWLGTGERINASGGAR"
 1914. 3371
 /gene="mlr3014"
 1914. 3371
 /gene="mlr3014"
 /codon_start=1
 /transl_table=11
 /product="homoserimidine synthase"
 /protein_id="BAB50002.1"
 /db_xref="GI:14023396"
 /translations="MLDKMANENWPPYGEITGVPMVIGFSGICRGTLPLIERHFKFDK
 SRNVLPDTRDKLDERGIAFQEAVENTNNYKLLTPLLTTGGGGQGFVNLSVDTG
 VSDTMCANRKLGVLYIDTVPEPWLGYFDKADNASTRTNYALREAMIKEKHDPGGAT
 VYTGCGANFGMSVFTYQKALVNLATDLGFESEPAQEDREGWAKMKKQKGVGHIAE
 RDTQRTYKPKPMDFWNTVSGVEGFISEGLOPAELGWGTHEKMPKMGKHKHKGSKAAI
 YLEOPCANTRVSRVCTPGCAQYGLLVTHNEAISIADPFTVRSKKGVQVYRPTCHYAH
 PNDAMLKLDMEFGAGKQPVHVLDELVDGDELGVLYGHDKNAYWVGSQSLSL
 AERKALPQDNATGMQVTSALAGWVALENPDAGIVEADMDYKRCCLDQVQSPYLGFLV
 KGYITDWTLPDRRPLGFLPFDLDRSDPQFNTILVR"
 3431. 3931
 /gene="mlr3016"
 3431. 3931
 /gene="mlr3016"
 /codon_start=1
 /transl_table=11
 /product="outer membrane lipoprotein"
 /protein_id="BAB50003.1"
 /db_xref="GI:14023397"
 /translations="MRPGAACVLRLDARPDQHFGRSHRAQNNRFLDRCVWRGFQMTI
 ARKVLISGWAGSFMMLAACTTSGPDAPPMAPAAKPGVGSWIDAKGTGLSTFTAGKFA
 TVATDTQKGLADGYSYTMGTGATSVETNGTSLIRQTPTVSNCLLIQTSQINLTSSGQNF
 VLTRRT"
 4123. 6243
 /gene="mlr3017"
 4123. 6243
 /gene="mlr3017"


```

LKNAIIIPDPYHGANEKAVQWVAQQDWIERTFILDDAEASWYLDIDYLDVIAIVFN
DVPVLSADNCFRYPEDISRAVPGENTIRIHFSNITAGAEQAQRPQFIPYHPGNS
PIANGNLRKPRQFGRWMDNIAIAPLGLYKILLKRLDARIIEHVVSOLHVPAGSVEQ
HVAWTLFAELGPASLPVYLSLGRLEKCGVAGETVVRHVFVFNPDLPWAPAGSEQ
TLVLTIVPEDEVTRQIGRTIELLTDKDEAGSREAFERLINGREIFCRGANWIPADAL
YSLTSREKTEDLLCSAVEANNNIRVWGGFYEDFYDLCDRLGLVWQDFACNL
YPCSEPLDNVEHVDYQVRLSSHSPIALWCNDNELVGLATWFDSSRNRRDRLVAY
DLNRNTEKAKATPEALWPPSPASGYLDYGDWADHAGDGMWYVSHENKSPDN
YHQVRFCEFOFQSVTSNPIRTVAEDKDMNIAFPVIELHOKNVGGNRIAGTWFR
YFQVRFENFVYSLOVQOALARTVAVYWSLPHCMGTLYWQLNDTWPVASWSSLD
RVPLKSNAGCTSDKKAATLTDIMDSILPDGAILLANPFIASNGMTGGHVRDRTYKALE
LQAPGLEFSVGPLUNGQFEIDVTAAGLALFIMLEADQOPGRYSNLPDLAAGETRRRIIF
TPKAGGQPHFRIFDLHTCOSSPNPGIETMRKA"
gene
5278. 6273
CDS
/gene="AGR_C_2811"
/gene="AGR_C_2811"
/notes="hypothetical protein b3001 - Escherichia coli
(strain K-12)"
/codon_start=1
/transl_table=11
/product="AGR_C_2811p"
/protein_id="AA87315.1"
/db_xref="GI:15156612"
/translation="MKYNHCGKTKLKLPAISLGLMHNFGNDTPHOTKQACIRRAFDLG
ITFDLANNYPGPGSAETAFGLKTDPRGYRDEMIISKAGYNNMPGPGYBWSRK
YLLSSCDQSLKRLGLDVIDFYSHRFDNPTPLEETCGALDQIVRSKGLVYGLSSVNS
KRTREAAILLKIDGTCLIHQPSYMINRWIEEDGLVDLIELIGISIVFSLAQML
TTYKLGVPDGRSASQSLNPAFLNENRINRALSIAERRGQTLAQMLAWLVRG
GRITSALIGASRVEQVEDCVKLDNAEFSTELAEITDRYAKDADINLWAKSAERV"
6465. 6791
gene
6465. 6791
CDS
/gene="AGR_C_2812"
/gene="AGR_C_2812"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2812p"
/protein_id="AA87316.1"
/db_xref="GI:15156613"
/translation="MKFVTIILAATVLAAPWAQASRHHDRHGVTVVERQVTTKKVI
VKHWRDQRGURASERRNNVDRYRHYHLAEPDRDQWRVVDNQFLINAVSGLIV
GLAAAR"
complement(6858..8288)
/gene="AGR_C_2814"
complement(6858..8288)
/gene="AGR_C_2814"
/notes="6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING"
/codon_start=1
/transl_table=11
/product="AGR_C_2814p"
/protein_id="AA87317.1"
/db_xref="GI:15156614"
/translation="MEQAEIGLIGLVMSGNLALNIAERKNIAVFNRTPEVTRKPYA
DAGELQGIIPCEITIEFVAARPPRIIMIKAGDPVQDQMEILKPLHANGDIMIDA
GNANPRDTRRFDNLKSLGTFTMGVSGEGRGHPGISIMVGGSTEDSKRVEKVLTS
ISAKYNDPDCVWLNGDAGHVKTHINGEYADQMIAEIIYGLRDLGKMSVAEIVL
FVANKERGLNSYLIBITEKTVRAADPTTGKEMVDLIDKACQKGTGKMSVLEAQMNG
VAAATELAAKSSQNDREAEKIFGLPTLAAAPADKFAITADLESALLAAKVG
AYAQGFPAISAKSEFNWNLNPMPTIARIWAGCIIRSEFLDEITSFTKDPHVNLIIV
TAFSAIVKDTDAPLRVVSVALSGLPVLSALASALGYFDAYRRGRGSANLIQAQRDF
FGAGFPTDGVDPKPGWGSADIP"
BASE COUNT 1814 a 2570 c 2348 g 1740 t
ORIGIN

Query Match 15.6%; Score 136.6; DB 1; Length 8472;
Best Local Similarity 56.8%; Pred. No. 3e-20;
Matches 312; Conservative 0; Mismatches 204; Indels 33; Gaps 2;

QY 325 TCAGTGGCTTCGCCCAATGTTGATCGCGCTCGTGGGAATCTCATCGATCGCTCC 384
|||||
Db 6263 TCAGCAGATTTTCCCGAGAGTTGATATCCGCATCTCTTGGCGTAAACGGTTCGATTCGGCC 6204
|||||

```

```

QY 385 AACTCGCGGTGAGAAATCCAAAGTTGTTGAGTGAATCAAGGCTGTTGCCAGTGCTCA 444
|||||
Db 6203 AGTCTCTCGTAGAGAACTCGGCATTTATCGATGCTTTTCAGCAGTCTCGACCTGTTGG 6144
|||||
QY 445 ACTGAGAAAGCACCATAATCAATCGTGTGTCACGGTATCCGCGCGCTACTCTCTCTGCTCG 504
|||||
Db 6143 ACACGGCTTGGCGCAATCAATGCTGAGGTAATGCGGCC-----GCCG 6102
|||||
QY 505 CGCAGACACCATGACAGCGCCATCTCGGCAAGTACTGCGCGGTTCTTGGGGGATGTCA 564
|||||
Db 6101 CGCAGAACCCAGCAATTCGCCATCTCGCGCCAGCTCTGGCCACGCGGCTCGGCAATGCTG 6042
|||||
QY 565 TTGAGCTTGGCGGACCATATCAATATTGTTTCAGTTCCAATCGCTTCAGAGAGCGCGGTGAGTGTG 5982
|||||
Db 6041 TTGAGCGCGCGGATGTTTTCGACATTCGCTGTTGAGAGAGCGCGGTGAGTGTGAGTGTG 5982
|||||
QY 625 CCTGCTCGCGCGGGAACCTCTGGAATTCATCGAGATATTGTTGCTCGAGCAGGCC 684
|||||
Db 5981 CTCTGTGAGGACCGGCTGCCATCGGCACACCGCCAGATATTTCGTCGTGAGCATGCC 5922
|||||
QY 685 TGGCGAAGTGTGAGAAACATGACGCCCAAGACCATTTGTTGGCAGCTGACTGCAACAG 744
|||||
Db 5921 TGGCGCAGCGCGGAAAGACGATGAGGCGGATACCCAGTTCTTCCAGCGTAT----- 5870
|||||
QY 745 TTCTCACCTGTCATCGCCCGGTTTCTCCACCCAGCATTAATGATGGAATAGCTTGGCTCA 804
|||||
Db 5869 -----CGACAGACCGTCTTCTCGATCCAGCGGTTGATCATGAATAGTCCGCTCG 5817
|||||
QY 805 TGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAACTCAGCGCGCTCCGCTGTGAGCTCT 864
|||||
Db 5816 TGGATGATCGAGGCGGTGCGGAGATCCTTCAGGATAGCGCGGCTCGCGGTGCGCTTC 5757
|||||
QY 865 GGACCGTAG 873
Db 5756 GAGTTGTAG 5748
|||||

RESULT 13
AE009111/13051 bp DNA linear BCT 20-DEC-2001
LOCUS
AE009111.1 Agrobacterium tumefaciens strain C58 circular chromosome, section
137 of 256 of the complete sequence.
ACCESSION AE009111 AE008688
VERSION AE009111.1 GI:17739946
SOURCE Agrobacterium tumefaciens str. C58 (U. Washington).
ORGANISM Agrobacterium tumefaciens str. C58 (U. Washington)
REFERENCE 1 (bases 1 to 13051)
AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,
Karp, P., Romero, P., Zhang, S., Yoo, H., Biddle, P., Jung, M.,
Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C.,
Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V.
and Nester, E.W.
The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
Science 294 (5550), 2317-2323 (2001)
JOURNAL MEDLINE
PUBMED 21608550
11743193
REFERENCE 2 (bases 1 to 13051)
AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,
Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,
Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,
Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,
McClelland, E., Palmieri, A., Raymond, C., Rouse, G.,
Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,

```

Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.

Direct Submission

Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357240, Seattle, WA 98195-7242, USA

FEATURES

Location/Qualifiers

```

1. .13051
  /organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
  /strain="C58"
  /db_xref="taxon:180835"
83. .1072
  /genes="znuA"
  /note="Atul521"
83. .1072
  /genes="znuA"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="ABC transporter, substrate binding protein [zinc]"
  /protein_id="AAL42526.1"
  /db_xref="GI:17739947"
  /translation="MKSILPLMASAALAAVAGATAADPVVVSIKPVHSLVAIMKG VGEPLIVGASPHYNYLRPSNARKLEKADVFWVPGLEAFLOKPLEALASKATVV ELDAKGLKLPKGGPEANDDGEHEAGHAGHTEDEGADHNDHAGSEHEHGA YTHLMDPANAKAAQALETALIDAGNAATYQNTKKGLDDLDALDAVEVETVKP EPQFPLVHDAYQVEHYGVKTAGSITVSPETLPGADRVKQMEKVRQLGAPCVFA EQFEPKLVITEGTAAKSATLIDPEATLTGPDILYFKLMRGIAGSLKNCLS"
  /complement (1094. .2113)
  /genes="Atul522"
  /note="Atul522"
  /complement (1094. .2113)
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="transcriptional regulator, LacI family"
  /protein_id="AAL42527.1"
  /db_xref="GI:17739948"
  /translation="MPSAMGGRTLTIAVTGLTITVSRLKADPADIGAEKERV LTAQIQYQNPAGVRLTKTNVIALVLSVDELMGFTSQMVFGITEVLSTQYHLV VTHIAKDSMPVIRVILETSADGVIISKIBNDPRVFRMTPFNPFVTHGRSDMGI EAHFDPDEAYEAVEALQCKRKRIAVIPPSRPSFDHARKGFNRGIDRGLTE FPDATVETPLEKIRDQRLMQSSDRPDGIVISGSSSTIALVAGFEAGVYKIGEDV DIVSKQSAEFLNWKIPQIHTVNEDIKLAGRELAKALLARINGEAETLSIGPVWVS MAPKP"
  /complement (5133)
  /genes="mana"
  /note="Atul523"
2650. .5133
  /genes="mana"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="beta-mannosidase precursor"
  /protein_id="AAL42528.1"
  /db_xref="GI:17739949"
  /translation="MISSSTPETVIDLAGLHLASVEGDHATEISIPGDIHSLKNAH IIPDHYHAKAVQWVAQODWIERTFILDDAASWYILDYLDVAVFYVNDVPLV SANCPRVYEDISRAVPCGENTIRHFNHSNITAGAEQARQPPYIPYHFGNSPIANG NMLRKQCHGMDWNTAIALPLGLYKILLKRLDARIEHVVSQHHVGGVELHVAVT LFAEGPASPLVLSGLDERLRCGCVAGETVVRHVFFENPDLPWAPGSGEOTLYKL TVELPDVETVQIGFTIELTLDKDEAGSRFAFRINGEIFPCGNAWIPADALYSLTS REKTEDLCSAVENANMMIRVWGCGFYEDFYDLCDLGLLVQODFMFACNLPQCE DFLDNDYHEVDYQVRLSHSPSIALWCQDNELGALTWFDESRNDRDYLVAVDRLNR TIEKALKATPEALWMPSSPASGLVDGDAWHADGSGDMHYSWVSHENKSFQNYHGVK PRCSEFGFGSYTSMPIRTYAECDKMNIASPVIELHQKNGVGNRIAGTMFYFRFP

```

gene

CDS

```

RDFNFVLSQVQALAIRTAVDWRSLSKPHCMGCTLYWQLNDTHPVASWSLDDGGGH
KALHYARRFPQVAVSAIPADSGRRVTFSMVNDTAEDVEIDMNIVALAMDGNRVLK
SANGTCSKRAATLTDIMDSLPDGAFLAWNFIASNGMTGEGHVRDITYKALEIOPAG
LEFVSGLKNGQFEIDVTAAGLALFIMLEADQPCGRYSNLFDLAAGETTRIIPTPKGA
GPQPHRIEDLHTCQSSPNPGIETMRKA"
5237. .6268
  /gene="Atul524"
  /note="Atul524"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="conserved hypothetical protein"
  /protein_id="AAL42529.1"
  /db_xref="GI:17739950"
  /translation="MVWQPAENRYASMKYNHCGTKGLKLPALISGLWHNFGNDTPHOT KOAICRRAPDLGITFDLANNYGPFGSAETAFCGLIKTDPRGYRDEMISSKAGYNN WPGYGEWGRKYLIISSCDQSLKRMGLDYVDIFYSHRPDPNTPLEETGALDQIVRSG KALVGISSYNSKRTREAAALIKDLGTPCIIHQPSYSMINRWIEEDGLVDLIEHLGIG SIVFSPLAQGLMTTKYLGSGVDPGSRASQSLNPAFLNRNVRNIALNSIAERRGQT LAQMAIAWLVLRGRITSALIGASRVEQVDCVKALDNAEFTSELAIEDRYAKDADIN LWAKSAERV"
6460. .6786
  /gene="Atul525"
6460. .6786
  /gene="Atul525"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="conserved hypothetical protein"
  /protein_id="AAL42530.1"
  /db_xref="GI:17739951"
  /translation="WKKFVTLLAATVLAAPMAQASRHHDRHGVTVVEROVTTKKVI VKXHRWRGQRLSARERNMVDYRRYHLAEPDRDQVRVVDNQFLINAVSLIV GLAAR"
  /complement (6853. .8283)
  /gene="gnd"
  /note="Atul526"
  /complement (6853. .8283)
  /gene="gnd"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="phosphoglucanate dehydrogenase, decarboxylating"
  /protein_id="AAL42531.1"
  /db_xref="GI:17739952"
  /translation="MEQAEIGLIGLVGMSNLALNIAEKGKNIATVFNTPVTRFYA DAGELQGGIIFCETIEFEVAARPPRIIIMKAGDPVDVQOQMEILKPLHANGDINDIDA GNANFRDTRIRFDNLKDSGLTFIGWVGSGEGAGHGSINVGGTDSWKRVERVLTS ISAKYNDPCVAVLNGDAGHFKVKTIHNGIEYADQMIAEYIGILRDGLKMSAVBIAD VFAEWNGRLNSYLIEITGEKVLRAADPTGKPMVDLIIDKAGQGTGKWSVEAQNMG VAATAEAARVAARILLSSQKDERAEKFGIPLTAAAPADKKAFTADESALLAAKVG AYAGFALMSAASKEFNWNLPMPTIARIWRACIIRSEFLDEISAFKDPHVAHLIV TPAFSAIVKDTADPLRRVVSALVSLGSLVPSALASALGYFDAYRRGRGSANLIQAORDF FGAHFERTDGVDPHGPWGSGADIF"
  /complement (8435. .8596)
  /gene="fixS"
  /note="Atul527"
  /complement (8435. .8596)
  /gene="fixS"
  /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
  /codon_start=1
  /transl_table=11
  /product="nitrogen fixation protein FixS"
  /protein_id="AAL42532.1"
  /db_xref="GI:17739953"
  /translation="MNMLTYLIPVALFLGALGLFALWMSVRSGQYEDMDGAARVIED GDRPRPSPS"
  /complement (8593. .10884)

```

gene

CDS

gene

CDS

gene


```

/ gene="fixi"
/ note="Atu1528"
complement (8593..10884)
/ gene="fixi"
/ note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/ trans1_table=1
/ product="nitrogen fixation protein Fixi"
/ protein_id="AAL42533.1"
/ db_xref="GI:17739954"
/ translation="MSCCARGTSGSLGELPPLNPPSEBLLASRLDGLQGLRDTLSV
PGVCGACITITIALGRIGOVERNLSSKRAVVMKEEVNGVTDPAIDARAILA
TGYRIHLFASGQDASVALRSQIRAVALCGFASANTMLLSVSVMSGADATRDMPHWI
SAMIAAPALTYGGRFFQYSAWALKHGRTNMDVPILALITLSYAVSLWETIHGHEHAW
FDATVSLFLLRDLHIMDKARSALAGLARLSPRGATVIDAGQVDRYPLADIE
PGMSIAAAGRDVADAVVCGSDLDMSLVNGESAPRRVWAGDSLQAGTLNLTGSIV
ARYTASAKDSFSEVIGLMEASGGRARYRIADRAASYSPVVHLLALTYFLGWGIF
GDWQKMLIAIAVLITPCPALGLVAVVQVVAAGRLFRGIMVKEGSAERLSRID
TVLFDKGTLTVGKPLRVETADVPKPAIMAIAGLAHRSRPLSKALHAAYNGALPAYE
AVREIPGSGVEATEAGTYRLGNRRFPACPDETRINNGDARSEVVLSDGRHLASFGFE

Query Match      15.68; Score 136.6; DB 1; Length 13051;
Best Local Similarity 56.8; Pred. No. 2.9e-20;
Matches 312; Conservative 0; Mismatches 204; Indels 33; Gaps 2;

QY 325 TCAGTGGCTTCGCCAAATGTCATCGCGCGTCGTGGGAAATCTCATCGATCGCTCC 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6258 TCGCAGATTTGCGCAGAGGTTCATATCCGCATCTTGGCGTAACGGTCATTCGGCC 6199
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 385 AACTCGCGCTCAGAAACTCCAGTGTGTGAGTGAATCAAGGCTGTGTTCAGCTGCTCA 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6198 AGCTCTCGGTAGAGAACTCGGCATTATCGAGTCTTTTCAGCAGTCTTCGACCTGTTCG 6139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 445 ACTGACGACGACCAATCAATCACTGCTGTCGATCGGTATCGCGCGCTACTCTCTTGTCTCG 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6138 ACACGGCTTCGCCAAATCAATGCTGAGGTAAATGCGGCC-----GCCG 6097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 505 CGCAGACCCATGCAAGCGCATCTCGCAAGTCACTCGCGGCTTCTCGGCGATGCTCA 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6096 CGCAGAACCCAGGCAATGCACTCGCGCGCTCTGCGCACCGCCCTCGCNAATGCTG 6037
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 565 TTGAGCTTGGCGACCATATCAATATTTGTCAGCTTCAACATGCCCTCAGACAGGCACTTA 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6036 TTCAGCGCGCGATGTTTTCGACATTCGCTCGTTGAGGAGCGCGGTTTCAGTGACTTG 5977
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 625 CCCTGGCTGGCGCGGACCCCTCTGGATTCATCGAGATATTTGTCCGTGAGCAGGCC 684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5976 CTCTGTGAGGACCGGCTGCCATCCGCGCACACCGCCAGATATTTCTGTCGTCATGCC 5917
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 685 TGCGCAAGTGTGAGAAGCAATGACCGCAAGACCATTTGTTGGCAGCTGACTGCAACAG 744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5916 TGCGCCAGCGCGAAGACGATGAGCGCATATCCCATGTTCTTCCACGCTAT----- 5865
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 745 TTCTACCGTCAATCGCCCGGTCTCTCCACCAACGATTAATGATGATAGTCTTGGCTGA 804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5864 -----CGACAGACCGTCTCTCTCGATCCAGCGTTGATCATCGATAGCTCGGCTGG 5812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 805 TGAATCAGACGGCGGAGCCCTCTCCGCCATGAATCTCAGCGGCTCCGCTGTGAGCTCT 864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5811 TGGATGATGACGAGCGGTGCGGAGATCTTTCAGGATAGCGCGGCTCGCGGCTGCGCTTC 5752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 865 GGACCGTAG 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5751 GAGTTGTAG 5743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 14

```

SC7H9/c      SC7H9      15586 bp      DNA      linear      BCT 12-MAY-2002
LOCUS      Streptomyces coelicolor cosmid 7H9.
DEFINITION      AL450223 AL645882
ACCESSION      AL450223.2 GI:20520842
VERSION

```

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

alanyl tRNA synthetase; amidase; integral membrane transport protein; katA2, catalase; oxidoreductase; regulatory protein; tetr-family transcriptional regulator; transcriptional regulator. Streptomyces coelicolor A3(2).

Streptomyces coelicolor A3(2)

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

1 (bases 1 to 15586)

Redenbach, M., Kiese, H.M., Denapalte, D., Eichner, A., Cullum, J., Kinashi, H. and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome

Mol. Microbiol. 21 (1), 77-96 (1996)

97000351

8843436

2 (bases 1 to 15586)

Seeger, K.J. and Harris, D.

Unpublished

3 (bases 1 to 15586)

Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Direct Submission

Submitted (16-NOV-2000) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK

On May 9, 2002 this sequence version replaced gi:1128487.

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics

Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web

(URI: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at <http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 7H9 lies between and overlaps with cosmids 5F1 and 2H2 on the AseI-A genomic restriction fragment.

FEATURES

source

```

1..15586
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 7H9"

```

misc_feature

```

2..113

```

```

/note="nominal overlap with Streptomyces coelicolor cosmid St7H9"

```

```

nominal overlap with Streptomyces coelicolor cosmid St5F1"
complement (583..618)
/note="possible stem loop. Score 51: 17/17 (100%) matches,
0 gaps"

```

stem_loop

```

gene
/complement(641..2161)
/gene="SC7H9.02c"
/note="SC07590; kataA2"
CDS
/complement(641..2161)
/gene="SC7H9.02c"
/note="SC7H9.02c; kataA2; catalase, len: 487 aa; similar to
SW:CAT_A_MTCUJ (D1R_S27264) Micrococcus luteus catalase (EC
1.11.1.6) Kata, 497 aa; fasta scores: opt: 1791 z-score:
2040.9 E(): 0; 54.8% identity in 484 aa overlap and to
SW:CAT_A_STRCO (EMBL:AL035478) Streptomyces coelicolor
catalase KATA or SC2G5.25c, 487 aa; fasta scores: opt:
1699 z-score: 1936.3 E(): 0; 53.8% identity in 483 aa
overlap. Contains Pfam match to entry PF00199 catalase,
Catalase and matches to Prosite entries PS00438 Catalase
proximal active site signature and PS00437 Catalase
proximal heme-ligand signature"
/codon_start=1
/transl_table=11
/product="catalase"
/protein_id="CAC16511.1"
/db_xref="GI:11228489"
/db_xref="SPTREMBL:Q9EWJ4"
/translations="MTDTPATTDGAPAESDEHSLTAGPAGVLLQDAYLIEQMAQ
FNERIPERQPHAKGSGAGHFVTHDVSAYTKAAVFPQGTDLVARESTVAGERGS
PDWRPRGFANKEVTSQGYNDWGNTPVFFVDMKQHFIRSKRRADNLRDHD
MQMDFWTLSPESAHQVTLMDGRGIPRTWRHNGYTSHTYMMINASGERFWKYHPT
DOGIEYFTHEADMAAADTDYHMDLFEHIRDGFPSTLHVQVMPEEAGYRFPN
FDLTKVMPHGIDYLPVGRMTLERNPTDHAETEQAAFOFNNLVPFGIPSPDRMLIAR
LFYSADAHYRIGANTQQLPVMAVVDVVRTYSKDGAMAYKNTDPPVAPNSKGGPAAD
TFHFGTPPSTWGTSTRTAYVSHAEDDWQGTAVREMDDAARDLVNDVNDHLL
NEVTEPVLQAFAYWNIDRTIGDRIAKGVRAKAGEKDFKAAEQRPARQAMQDKA"
/complement(719..2137)
/gene="SC7H9.02c"
/note="Pfam match to entry PF00199 catalase, Catalase,
score 925.90, E-value 1.1e-274"
/complement(1136..1162)
/gene="SC7H9.02c"
/note="PS00437 Catalase proximal heme-ligand signature"
/misc_feature
/complement(1979..2029)
/gene="SC7H9.02c"
/note="PS00438 Catalase proximal active site signature"
2217..2632
/note="High G+C content region (80.52%)"
/gene="SC7H9.03c"
/complement(2636..3598)
/note="SC07591"
/note="SC07591"
/complement(2636..3598)
/gene="SC7H9.03c"
/note="SC7H9.03c; probable regulatory protein, len: 320
aa; similar to SW:NAHR_PSEPU (EMBL:J04233) Pseudomonas
putida transcriptional activator protein Nahr, 300 aa;
fasta scores: opt: 452 z-score: 510.2 E(): 7e-21; 32.4%
identity in 296 aa overlap. Contains Pfam matches to
entries PF00126 HTH 1, Bacterial regulatory
helix-turn-helix protein, lysR family and PF01046
NodD_C term, NodD transcription activator carboxyl
terminal region and match to Prosite entry PS00044
Bacterial regulatory proteins, lysR family signature"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAC16512.1"
/db_xref="GI:11228490"
/db_xref="SPTREMBL:Q9EWJ3"
/translations="MNSTDPTVIDANLVALDIALAEQSVTRAAARMRTSPAA MSR
TLGRRLLDQPLLVRAGQVTPTRALARDDEATVVRIGALLTPATGVDPATLRS
TFSLOTDIALVGAALGLLTLAGEAPGVSLRFARELEAGPALRGVLDVGAIDH
VDRETRVELVTLMTAAVAPGHPLEGTPTAARPAARAAHVGVSRGRDGPVDALIA
EHGSRKRVSVTLGHLLAALSLAARTDVVALVPAARDEADFPSPFLTEQARVLGICLID
IPLPLPPTVITGMAWHPHRTADGAHMLRAAVRVRVLRAPPSAREG"
/complement(2849..3079)
/gene="SC7H9.03c"
/note="Pfam match to entry PF01046 NodD_C term, NodD

```

```

transcription activator carboxyl terminal region, score
37.90, E-value 3.9e-10"
/complement(3134..3562)
/gene="SC7H9.03c"
/note="Pfam match to entry PF00126 HTH 1, Bacterial
regulatory helix-turn-helix protein, lysR family, score
61.20, E-value 1.7e-14"
/complement(3428..3520)
/gene="SC7H9.03c"
/note="PS00044 Bacterial regulatory proteins, lysR family
signature"
3702..3706
3711..4604
/gene="SC7H9.04"
/note="SC07592"
3711..4604
/gene="SC7H9.04"
/note="SC7H9.04; hypothetical protein, len: 297 aa;
similar to TR:Q9RCY4 (EMBL:AL133469) Streptomyces
coelicolor hypothetical 31.4 kDa protein SCM10.14c, 299
aa; fasta scores: opt: 791 z-score: 890.7 E(): 0; 46.4%
identity in 302 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAC16513.1"
/db_xref="GI:11228491"
/db_xref="SPTREMBL:Q9EWJ2"
/translations="MIVTAPITGNIRLLPLLESAPARGEELRVIVRDPARLAAPV
RERVVTSHGDPVAVDRAFDGADAVFMLVPPDSRTPDDAYRGFTPAARALVAHG
VGHVGVSAIGRGTPVARAGLVTASLAVDOLLAETGVAREALASPSFENILEADS
IREHCFDVLVDADKSPVAVADIASVAGLLLDSDRWGTGTSVPVPLGPRDLSFNDLA
RIMERLGRPVYERHTDDLRSLVGLDKDFVDGVDMKRAKDDGLDAGVTRTPD
TATPTTFEDVWARTLGPVLS"
4679..4682
4688..5053
/gene="SC7H9.05"
/note="SC07593"
4688..5053
/gene="SC7H9.05"
/note="SC7H9.05, unknown, len: 121 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein SC7H9.05"
/protein_id="CAC16514.1"
/db_xref="GI:11228492"
/db_xref="SPTREMBL:Q9EWJ1"
/translations="MPESTATPTVTFMLVKTPEWALTVEERVHAFTEVVPVRA
RTAGVRSRFPYTFEYFYSARVTDVWWEADDDHAYQLLVDALETPFWDYRFEVVDLAVG
TENGVARYGVDAVTTLST"
Query Match 15.3%; Score 133.6; DB 1; Length 15586;
Best Local Similarity 57.5%; Pred. No. 1.4e-19;
Matches 300; Conservative 0; Mismatches 189; Indels 33; Gaps 2;
QY 352 CCGGCGTGTGGGAATCTCATCGATCGCCTCAACTCGCGTCAGAAACTCAAGTTG 411
|||||
Db 6218 CCGTCGACGACCGCGTACTCGTATCGGCCAGTTCGTCGTCGAGGACGGGTTG 6159
|||||
QY 412 TTGAGTGAAATCAAGCGTGTTCACGCTGCTCAACTGACGAGCAACCAATCAATGACTG 471
|||||
Db 6158 GCAGAGCGCGACGCTCTCTCCAGCTGCTCGGCCGGGAGGACCGATCACAGGAG 6099
|||||
QY 472 GTCACGGTATCCGCGCGGTACTCTCTGCTCGCGAGCAGCACCATGCAAGCGCATCTGC 531
|||||
Db 6098 GTCAAG-----CGCGGTGCGCAGCGCCCGCAGCGAGCGCATCTGC 6057
|||||
QY 532 GCAAGTCACTGCCCGGTTCTCTGGCGATGTCATTGAGCTTGGGACCATATCAATTG 591
|||||
Db 6056 GCAGAGCTCTGGCGCGCGCGCGATGTCGTTGAGGCGCGCAGCGCGGACGGTG 5997
|||||
QY 592 TTCACGTTCAACATGCCCTCAGACAGGACTTACCTCGCTGGCGGGGAGAACCTCTCGGA 651
|||||

```



```

predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE EXODEOXYRIBONUCLEASE III PROTEIN"
/protein_id="CAC46228.1"
/db_xref="GI:15074583"
/db_xref="SPTREMBL:Q92PT1"
/translations="MKIATYNVNGRGLRVLLRWLEASPDVVCVQLQELKAPDKFPV
KATLAAGYGAIMHGQKSWNGVAILARDREPTLRKGLPGDDPDTHSRYIEAAVEGMVI
GCILPNGNPGPKPEYKELPHRLTAVAAELLELDVPIVLADGYNVMPTELDVYKP
ERVNDALFRIEVRDVAHRLLEQSWTDALRQLRGPVTVTFWDFENAFARDAGLRID
HLLSPHVLRLSAGVDRHVRGWEHTSDHAPWIELSDGPAEDQ"
5519. 5737
/gene="SMC00958"
CDS
5519. 5737
/gene="SMC00958"
/functions="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAC46229.1"
/db_xref="GI:15074584"
/db_xref="SPTREMBL:Q92PT0"
/translations="MTWMDWRAELRLLEKLVDAVVAGARQDVVAAIILDEIGQLRL
AYBRDPDADIIEQVIEEPSNNWPAAD"
5819. 6005
/notes="Sm-4 OR SMC04555
REPEAT SM-4
predicted by Homology"
/evidence-not_experimental
/complement(6137..9652)
/gene="mfd OR SMC04450"
/complement(6137..9652)
/gene="mfd OR SMC04450"
/functions="macromolecule metabolism; macromolecule
synthesis, modification; dna - replication, repair,
restr./modif"
/notes="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE TRANSCRIPTION-REPAIR COUPLING FACTOR
(TRCF) PROTEIN"
/protein_id="CAC46230.1"
/db_xref="GI:15074585"
/db_xref="SPTREMBL:Q92P89"
/translations="MTMIPGDERKILAAITREITIGVPVPTGAEALVLAEIARAGAPVA
YILSDQKQVADLQVGFVAPDIPVLTLPQWDCLPYDRVSPSADTSARRLAALSALIA
HRRKHPAIVLVTINAAQLRISPDVIESLAFTAPFQNGQIRMDDLAARLERNFERVP
TVRMGEFVARGGILDVYVPGSGEPLRLDPFDGTLAIRSFDPASORTIGQVRSLDLN
PMSEVSLTPTTIGHFRKOYLSLFGAATRDDALYQVSEGRRYAGMEHWLPLFYDRLET
VFYLDGFRIVTDHIAEAAERSKLVDYDARLASASPKSQVTTGGTPYKVPDPM
LYLTAKGFEALNDLNAVLSPTEHGEARQVNNIARQGLRWAPAGEAUNDGRTI
NVFDQVKAHAKKARKAVIVSGWTEGSLDRLLQVLAEHGLANIRPVKALSIGSLK
PGEAASAVILEAGFETGLDVIWIGQDILGDLRVRRSKRRKRGADFAEVTGLDEGSY
VVAHAGIGFVGLRTTEAAGAPHDCLLELVADDAXLFLPVENIELLSRYSGSDAV
LDKLGVAWAKAKRLKLLDMAGGLIRIAERHHTHA PVLAAQDGYVDEPAARFPY
DETEDQLNSIDAVRDDLRGRPMWDLVCGVGEKTEVALRAAFIAAMGVQVAVVP
TTLARQKFTFSDRFGLRIRIQQASRUVGSKDALTAKKEVAEGKTDIVVGTHALLG
SSIKFANLGLLIDEEQHFVGKHKELKELTDVHVLTSATPIPTLQLALTGVREL
SLITTPVDMAVNTFISPFDAVIRFETLMREHYRGQSFCVPCVFDLPFIHDFLKS
DVPELVAVAHQMPATELEDIMNAFVEGYDVLLSTTVIESGLDVPNTANTLIVHRA
MFGLAQLYLRGRVGRSKRAFALFTLPVKNKLTGPAERRLKVLQSLDITLIGAGFQLAS

```

```

HDLDIRGAGNLLGEEQSGHIKEVGFELYQOMLEEAVALKGEERITHDQMSPOISVGT
PVMIFEEYVPDNLNRLGLYRRLGELTDLKEIDGFAELIDRFGLPFTVEVQHLKIVYV

Query Match      13.2%; Score 115; DB 1; Length 306250;
Best Local Similarity 53.6%; Pred. No. 1.8e-15;
Matches 309; Conservative 0; Mismatches 235; Indels 33; Gaps 2;

QY 286 AAACATGATGTTGATGGTTAGTTTTCGCGGGTTTGGAAATCGGTGCGCCCTTCCGCGCAAATG 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297332 AAATCAGGCAAAATCAATGATTCAGAGCAATTCAGAGCTCGCGCGATTTGGCCCGAGAGA 297391
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 TTGATCGCGCGCGTGGGGGAAATCTCATCGATCGCTCCAACTCGCGCGTCAGAAACTCC 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297392 TTGACGTGCGCCCTCGCGGGCGTAAAGATCGATCTCGCTCAGTTCTCTCGTGTGAAGTG 297451
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 AAGTTGTTGAGTGAATCAAGGGCTGTTGTCCAGCTGCTCAACTGACGAAGCACCAATCAAT 465
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297452 TCGTTTTCAGCGCCTTGACGAGTCGACGATCTGTGCGGAACGGCTGGCGCGATCAGC 297511
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 GCATCTGTCACGGTATCGCGGCGGTACTCTCTTGTCTCGCGCAGCACCCTATGCAAGCGCC 525
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297512 GCCGAGGTGATGCGCGCGC-----CGCGCAGCACCCAGCGCGAGTGGC 297553
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 526 ATCTGCGCAAGTGACTGCGCGGCTTCTTGGCGGATGTCATTGAGCTTTCGGGACCATATCA 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297554 AUTCAGCGAGGCTCTGGCGCGCGCTTTCGGCGATGGCGTTTTCAGCTTGGGATATTGTG 297613
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 586 ATATTGTTTCACTGTTCAACATGCGCTCAGACAGGAGCTTACCCCTGGCTGGCGGGAAACCC 645
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297614 ATGATCCCGGACGGATGAAGTCTTCTTGAGGAAGTGGTTCTGCGCGCGCGGTGTCC 297673
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 646 TCTGGAATTCATCGAGATATTGTCGCGAGCAGCGCCCTCGCGCAAGTGGTGAGAAAGCA 705
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297674 TCGGGAATGCGGTTTCAGGTAATTCGTCGTCAGCATCCCTCGCGCAGCGCGCGAGAG 297733
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 706 ATGACCCCAAGACCATTTGTTGGCAGCTGACTGCAACAAGTTCTCACCGTCATCGCCCGGT 765
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297734 ATCGAGCCGACGCGCGAGTTCTCTCACCGTGTGACGAGACCGTCA----- 297778
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 TCCTCCACCCCAACGATTAATGATGGAATAGCTTGGCTGATGATCAAGAGCGGCGAGCC 825
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297779 TCCTCGACCCCAACGGTTGAGCATCGAATAGTCGCGCTGGTGGATGAGACAGCGCGT 297838
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 826 TCCTCGCCCATGAACTCAGCGCGCTCCGCTGTGAGCT 862
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297839 ATCTCTTCAGGATTACCGCGCGCTCGCGGTAACGCT 297875
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: April 26, 2003, 23:44:41

Job time : 3110.77 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:02:03 ; Search time 1163.26 Seconds
(without alignments)
12154.292 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1421_2293

Perfect score: 873

Sequence: 1 gtcgataagcagcgggttg.....ctgtgagctctgaccgtag 873

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_etc:*
9: gb_est1:*
10: gb_est2:*
11: gb_etc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	56.4	6.5	860	17	AF075981
C 2	48.8	5.6	1101	17	CNS017SY
C 3	44.4	5.1	664	17	CNS03K4J
C 4	43.8	5.0	3237	17	BH770958
C 5	41.6	4.8	558	14	BM712036
C 6	41	4.7	617	14	BM729055

C 7	41	4.7	770	12	BE898495
C 8	41	4.7	861	12	BE740880
C 9	40	4.6	357	9	AI418751
C 10	40	4.6	431	9	AA452858
C 11	40	4.6	447	14	W91005
C 12	40	4.6	463	9	AA476500
C 13	40	4.6	504	9	AI159901
C 14	40	4.6	540	9	AI401605
C 15	40	4.6	564	9	AI820014
C 16	40	4.6	600	12	BE806166
C 17	39.8	4.6	1101	17	CNS00KK2
C 18	39.6	4.5	925	17	CNS0091P
C 19	39.2	4.5	846	17	CNS010RJ
C 20	38.6	4.4	436	10	BE025218
C 21	38.6	4.4	513	14	BQ818660
C 22	38.6	4.4	828	17	CNS026XT
C 23	38.6	4.4	830	17	CNS01MQH
C 24	38.6	4.4	1667	11	AY108197
C 25	38.4	4.4	650	10	BB576857
C 26	38	4.4	212	9	AA150978
C 27	38	4.4	486	9	AA149286
C 28	37.8	4.3	804	17	CNS01HIJ
C 29	37.4	4.3	707	10	AW190228
C 30	37.4	4.3	724	13	BI808515
C 31	37.4	4.3	731	14	BQ578715
C 32	37.2	4.3	448	17	BH614401
C 33	37.2	4.3	451	17	BH843100
C 34	37.2	4.3	1623	17	BH614228
C 35	36.8	4.2	312	10	AW015929
C 36	36.8	4.2	482	12	BG048617
C 37	36.8	4.2	506	13	BI683498
C 38	36.8	4.2	552	10	AW659102
C 39	36.8	4.2	692	17	CNS007WH
C 40	36.6	4.2	636	13	BI724148
C 41	36.6	4.2	819	12	BF064683
C 42	36.6	4.2	837	13	BI868875
C 43	36.4	4.2	660	13	BI328066
C 44	36.4	4.2	910	17	CNS006ON
C 45	36.4	4.2	934	12	BF311804

ALIGNMENTS

RESULT 1
LOCUS AF075981/c
DEFINITION AF075981 Salmonella typhimurium LT2, Lambda DASH II Salmonella typhimurium genomic clone 390-T3, DNA sequence.
ACCESSION AF075981
VERSION AF075981.1 GI:3320851
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
REFERENCE 1 (bases 1 to 860)
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library: comparison to the Escherichia coli K12 genome
JOURNAL PEMS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE 99243757
COMMENT Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifsci.sdsu.edu
Class: shotgun.
Location/Qualifiers
1..860
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"

modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-DWI is a normalized cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7f3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTCGGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

113 a	169 c	185 g	90 t	1 others
-------	-------	-------	------	----------

BASE COUNT

BASE COUNT	113 a	169 c	185 g	90 t	1 others
ORIGIN					
Query Match	4.8%;	Score 41.6;	DB 14;	Length 558;	
Best Local Similarity	46.9%;	Pred. No. 1.2;			
Matches 128;	Conservative	0;	Mismatches 145;	Indels 0;	Gaps 0;
Qy	98	CGGGCTCGGGCGCAATA	CGGCGCACCGGACGGTGGATTTTCCGCTGGCGGGTTCG	157	
Db	532	GCAGCGCCAGCGCGCGAACTCTCCACGTCCGCGAACCTGGCGCTCTCGCGGTTTCA	473		
Qy	158	CGGCAAGCCTCATCTGGTTTCCGCTTGGTGGGTTTCGGCGCAGCAGCATTTGTACGCCCGC	217		
Db	472	CCACGCTGTGGACTGCGACACCGTCACGTTGATGCCGTTCTCCAGCGCTTCGGCGCGC	413		
Qy	218	TGTCAGCCCCCAAGGTGTGGCGCTGGATCAACGTGCTGTGGCAGTTGTGTATGACCGGAT	277		
Db	412	TGGTCAACCGCAGCGGTGNCAGGTCGCTCTCGGGCGTGTGGTCTTTGATGAAGTAGTCGG	353		
Qy	278	TGGCCATCAAACTGATGTTTGATGGGTTAGTTTTCGCGGTTTTCGGAACTCGGTGGCCTTCG	337		
Db	352	TGTCCTCCCTCCGATGGTGAAGTCAGGTTCTCCAGGTAGAACGCGTTGTTGACACGG	293		
Qy	338	CCCAAAATTTGATGCGCGCGTCGTGGGAAATCT	370		
Db	292	CCGCGACCTTGATGCAGTCTCTGTTGGCGGATGT	260		

RESULT 6
BM729055/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

BM729055 617 bp mRNA linear EST 01-MAR-2002
UI-E-R01-aiw-a-06-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone
UI-E-R01-aiw-a-06-0-UI 5', mRNA sequence.
BM729055
BM729055.1 GI:19050388
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 617)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).

<p>The following repetitive elements were found in this CDNA sequence: 148-171, >GC-rich#low_complexity Seq primer: M13 Reverse.</p>					
FEATURES	Location/Qualifiers				
source	1..617				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="UI-E-EOI-a1w-a-06-0-UI"				
	/clone_lib="UI-E-EOI"				
	/tissue_type="fetal eye"				
	/dev_stage="fetal"				
	/lab_host="DH10B (Life Technologies) (Tl phage resistant)"				
	/note="Organ: eye; Vector: pVT3-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; UI-E-EOI is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pVT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGGTATACC. This library was created for the program, Gene Discovery in the Visual system, supported by National Eye Institute (NEI)."				
BASE COUNT	124 a	189 c	211 g	93 t	
ORIGIN					
	Query Match 4.7%; Score.41; DB 14; Length 617;				
	Best Local Similarity 46.9%; Pred.No.1.9;				
	Matches	128;	Conservative	0;	Mismatches 145; Indels 0; Gaps 0;
Qy	98	GCGCGTCGGGCGCAATACGGCGACACCGGACGGTGAGATTTCGCCGCTGGCGCGTTGG	157		
Db	537	GCAGCGCAGCGCGCCGAACCTGCATCTCCAAGTCCGCGAACCTGCGGCTCTGCGGTTCA	478		
Qy	158	CGCAACCGCTGATCTGGTTCGCCGTGCTGGGTTCGGCGCAGCAGCATTTGCACGCCCGC	217		
Db	477	CCACCGTGGTAGCTCCGACACCGTCACGTTGATGCCGTTCTCCAGCGCCTTGGCGCGCG	418		
Qy	218	TGTCAGCCCCAAGTGTGGCGCTGGAGTCAACAGTCGCTGGCGAGTTGTGATGACCGCAT	277		
Db	417	TGGTCAACCGCAGCGTGCCACAGTCGCTCTCGGCGTGTGTGTTTCGATGAGTAGTGGC	358		
Qy	278	TGGCCATCAAACCTGATGTTGATGGGTTAGTTTTTCGCGGGTTTTTGGAAATCGGTGCCTTCG	337		
Db	357	TGTCCTTGCCCTCGATGTTGTAAGTGACAGTCTTCCAGGTAGAACGGCTGTTGTGACACGG	298		
Qy	338	CCCAATATTGATCCCGCGCTCGTGGGAATCT	370		
Db	297	CCGCGACCTTGATGACTGCTCTCGTTGGCGATGT	265		
RESULT 7					
BE898495/c					
LOCUS	BE898495 770 bp mRNA linear EST 29-SEP-2000				
DEFINITION	601681396F1 NIH_MGC_9 Homo sapiens CDNA clone IMAGE:3951622 5', mRNA sequence.				
ACCESSION	BE898495				
VERSION	BE898495.1 GI:10365029				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 770)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/.				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D.				

Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLC819 row: p column: 23
 High quality sequence stop: 765.

FEATURES

Location/Qualifiers
 1. 770
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3951622"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 147 a 245 c 277 g 101 t
 ORIGIN

Query Match 4.7%; Score 41; DB 12; Length 770;
 Best Local Similarity 46.9%; Pred. No. 2.2;
 Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
 QY 98 GCGCGTGGCGGCAATACGGCAGACCGGAGTGGATTTTCGCGCTGGCGCTTCG 157
 |||||
 Db 385 GCACGGCAGCGCGCGAATGATCTCCAGTCGGCGAACCTGCGCTCGCGCTTCA 326
 QY 158 CGGAAGCTGATGTTCCCGTGGTGGTTCGCGCAGCAGCATTTGTCACGCCCGC 217
 |||||
 Db 325 CCACCGTGGTGGATCGCAGACACCGCTGATGTCGCTTCACGCGCTTCGCGCGC 266
 QY 218 TGTCCAGCCCAAGTGGCGCTGGATCAACGTCGTCGGGAGTGTGATGACCGCAT 277
 |||||
 Db 265 TGTCAACCGCAGCGTCCCGAGTCGCTCTCGGCGCTGGTGGTCTTGATGAAGTAGTGG 206
 QY 278 TGGCCATCAAACTGATGTTGATGGTGTAGTTTTCGCGGGTTTGGATCGTGGCTTCG 337
 |||||
 Db 205 TGTCTTGGCTCGATGGTGAAGTCAGGTTCTCCAGTAGAAGCGCTGTTGAGCAGG 146
 QY 338 CCCAAATGTTGATCGCGCGCTCGTGGGAAATCT 370
 |||||
 Db 145 CCGCCACTTGTGATGTCAGTCTCTGTTGGCGATGT 113

RESULT 8
 BE740880/c
 LOCUS
 DEFINITION 601593311F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3947138 5', mRNA linear EST 15-SBP-2000
 mRNA sequence.
 ACCESSION BE740880
 VERSION BE740880
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 861)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLC808 row: f column: 03
 High quality sequence stop: 754.

FEATURES

Location/Qualifiers
 1. 861
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3947138"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site: 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 187 a 245 c 285 g 144 t
 ORIGIN

Query Match 4.7%; Score 41; DB 12; Length 861;
 Best Local Similarity 46.9%; Pred. No. 2.3;
 Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
 QY 98 GCGCGTGGCGGCAATACGGCAGACCGGAGTGGATTTTCGCGCTGGCGCTTCG 157
 |||||
 Db 675 GCACGGCAGCGCGCGAATGATCTCCAGTCGGCGAACCTGCGCTCGCGCTTCA 616
 QY 158 CGGAAGCTGATGTTCCCGTGGTGGTTCGCGCAGCAGCATTTGTCACGCCCGC 217
 |||||
 Db 615 CCACCGTGGTGGATCGCAGACACCGCTGATGTCGCTTCACGCGCTTCGCGCGC 556
 QY 218 TGTCCAGCCCAAGTGGCGCTGGATCAACGTCGTCGGGAGTGTGATGACCGCAT 277
 |||||
 Db 555 TGTCAACCGCAGCGTCCCGAGTCGCTCTCGGCGCTGGTGGTCTTGATGAAGTAGTGG 496
 QY 278 TGGCCATCAAACTGATGTTGATGGTGTAGTTTTCGCGGGTTTGGAAATCGTGGCTTCG 337
 |||||
 Db 495 TGTCTTGGCTCGATGGTGAAGTCAGGTTCTCCAGTAGAAGCGCTGTTGAGCAGG 436
 QY 338 CCCAAATGTTGATCGCGCGCTCGTGGGAAATCT 370
 |||||
 Db 435 CCGCCACTTGTGATGTCAGTCTCTGTTGGCGATGT 403

RESULT 9
 AI418751
 LOCUS
 DEFINITION tg33a02.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:2110538 3', mRNA linear EST 30-MAR-1999
 mRNA sequence.
 ACCESSION AI418751
 VERSION AI418751.1 GI:4264682
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 357)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.


```

/db_xref="taxon:9606"
/clone="IMAGE:2103625"
/clone_lib="NCI CGAP Brn23"
/tissue_type="Glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAAGTGGAGCGCGGCATCTTTTTTTTTTTTTTTTTTTT T 3'); double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      103 a 154 c 159 g 124 t
ORIGIN

Query Match      4.6%; Score 40; DB 9; Length 540;
Best Local Similarity 53.1%; Pred.No.3.4;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps. 0;

QY 211 CGCGCGCTGTCAGACCCAGGTGTGGCGCTGGATCAAGTCTGCTGCGAGTTGTGATG 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 CGCGCGCTGTGTCAACGCAGCGTGCACAGTCTGCGCGCGTGTGTCATTGATGAAG 61

QY 271 ACCGATTGGCCATCAAACTGATGTGTATGGGTGTAGTTTTTCGCGCGTTTTTGGAAATCGGTG 330
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 TAGTGGGTGTCTTGGCCCTCGATGTGTGAAGTGCAGTTCTCCAGGTAGAAGCGGTGTG 121

QY 331 GCCTTCGCCCAAAATGTGTATGTCGCCGCGTCTGTGGAAATCT 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 AGCAGCGCGCCACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 161

RESULT 15
AI820014
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AI820014          564 bp      mRNA      linear      EST 21-DEC-1999
wJ60d07.x1 NCI CGAP Lu19 Homo sapiens cDNA clone IMAGE:2407213 3'
similar to TR:070465 O70465 D0C4.; mRNA sequence.

AI820014
AI820014.1 GI:5439093
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
1 (bases 1 to 564)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Prepared by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 859 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 498.
Location/Qualifiers
1. .564
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2407213"
/clone_lib="NCI CGAP Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"

```

```

/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT7T3 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."

```

BASE COUNT 113 a 158 c 164 g 129 t
ORIGIN

```

Query Match 4.6%; Score 40; DB 9; Length 564;
Best Local Similarity 53.1%; Pred. No. 3.5;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 211 CGCCCGCTGTCCAGCCCGGCGGTGGCGTGGATCAACGTCGTCGTCGCGAGTTGTGATG 270
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 2 CGCCCGCTGTGTCACCCGCGCGTCCCGAGGTCGCTCTCGGGCGTGGTCTTGATGAAG 61

QY 271 ACCGCATTGGCCATCAAACTGATGTTGATGGTTAGTTTTCGCGGTTTTCGATCGGTG 330
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 62 TAGTGGGTGTCCTTGGCCCTCGATGGTGAAGTGCAGGTTCTCCAGGTAGAAGGCGTTGTTG 121

QY 331 GCCTTCGCCCAAAATGTTGATCCCGCGCTCGTGGGAAATCT 370
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 122 AGCAGCGCGCCACCTTGATGTCAGTCTCTGTTGGCGATGT 161

```

Search completed: April 27, 2003, 00:51:28
Job time : 1172.26 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 17:06:23 ; Search time 192.113 Seconds
(without alignments)
10233.535 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1421_2293

Perfect score: 873
Sequence: 1 gtcgataagcagcggtttg.....ctgtgagctgtgacgctag 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	873	100.0	2374	18	AA796816
2	873	100.0	349980	22	AAH68528
3	568	65.1	627	22	AAH68419
4	404.2	46.3	993	22	AAF68077
5	326	37.3	822	22	AAF71779
6	326	37.3	822	23	AA596098
7	303	34.7	708	22	AAH68420
8	193.6	22.2	1568	22	AAH45375
9	71.6	8.2	5541	15	AAQ55755

C 10	65.4	7.5	988	24	ABK72786
C 11	57.6	6.6	23128	23	AAS59552
C 12	57	6.5	15239	17	AAT33536
C 13	57	6.5	4403765	22	AA199683
C 14	57	6.5	4411529	22	AA199682
C 15	56.4	6.5	636	21	AAA52691
C 16	52.6	6.0	2365589	24	ABA30521
C 17	50.2	5.8	1277	24	AAD31869
C 18	41	4.7	2157	22	AAH14096
C 19	41	4.7	3270	22	AAH14671
C 20	41	4.7	3614	22	AAH14183
C 21	40	4.6	399	24	AAS20981
C 22	40	4.6	431	24	ABL65724
C 23	39.6	4.5	1935	16	AAQ92069
C 24	38.2	4.4	948	22	AAH66637
C 25	38.2	4.4	1059	22	AAF71735
C 26	38.2	4.4	1059	22	AAF72186
C 27	38.2	4.4	349980	22	AAH68529
C 28	37.8	4.3	109519	22	AAS08693
C 29	37.6	4.3	4403765	22	AA199683
C 30	37.6	4.3	4411529	22	AA199682
C 31	37.4	4.3	1139	9	AA80100
C 32	37	4.2	1155	22	AAF26425
C 33	37	4.2	3870	23	ABL28047
C 34	37	4.2	6556	23	ABL28046
C 35	36	4.1	1068	22	AAH65404
C 36	36	4.1	349980	22	AAH68525
C 37	35.8	4.1	791	24	ABK34316
C 38	35.8	4.1	956	13	AAQ20267
C 39	35.6	4.1	1634	24	ABN85320
C 40	35.4	4.1	996	23	AAS4138
C 41	35.2	4.0	530	21	AAC74606
C 42	35.2	4.0	530	24	ABN18439
C 43	35.2	4.0	1815	22	AAC44497
C 44	35.2	4.0	12152	22	AAS08699
C 45	34.8	4.0	707	22	AAI64001

ALIGNMENTS

RESULT 1

AA796816

ID AA796816 standard; DNA; 2374 BP.

XX AA796816;

AC AA796816;

XX 12-MAR-1998 (first entry)

DT DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.

DE LysG, LysE; ORF3; lysine transport; regulatory protein; export protein;

XX Microbial production; amino acid; animal feed additive; ds.

KW *Corynebacterium glutamicum*.

OS *Corynebacterium glutamicum*.

XX Key Location/Qualifiers

FT CDS complement (82..954)

FT /*tag= a

FT /label= LysG

FT 1016..1726

FT /*tag= b

FT /label= LysE

FT complement (1723..2373)

FT /*tag= c

FT /label= orf3

XX DE19548222-1.

PN *Corynebacterium glutamicum*

XX *Corynebacterium glutamicum*

XX *C. glutamicum* gene

XX *C. glutamicum* codin

XX *C. thermosaminogene*

XX *Escherichia coli* g

XX 26-JUN-1997.

XX 22-DEC-1995; 95DE-1048222.

PR 22-DEC-1995; 95DE-1048222.
 XX (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 XX Eggeling L, Sahm H, Vrljic M;
 PI
 XX WPI; 1997-333867/31.
 DR P-PSDB; AAW37714-16.
 XX
 XX Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT Corynebacterium
 XX
 PS Claim 23 and 26; Page -; 16pp; German.
 XX
 CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
 CC LysE and ORE3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine, as
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.
 CC NB: This sequence has been created from the information given in table 2
 CC of the specification.
 XX
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

Query Match 100.0%; Score 873; DB 18; Length 2374;
 Best Local Similarity 100.0%; Pred. No. 7.3e-247;
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCCATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGCAATCGTCACTGGTGGT 60
 DB 1421 GTCCATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGCAATCGTCACTGGTGGT 1480
 QY 61 AACCCGAATGCGTATTTGGAGCGGTTTGTGTTATCGCGCGCGTCCGCGCAATACGGC 120
 DB 1481 AACCCGAATGCGTATTTGGAGCGGTTTGTGTTATCGCGCGCGTCCGCGCAATACGGC 1540
 QY 121 GACACCGAGCGGTGGATTTTCGCGCTGCGCGTTCGCGGCAAGCCTGATCTGGTCCCG 180
 DB 1541 GACACCGAGCGGTGGATTTTCGCGCTGCGCGTTCGCGGCAAGCCTGATCTGGTCCCG 1600
 QY 181 CTGGTGGGTTTCGCGCAGCAGCATTTGTCAGCCGCTGTCCAGCCCAAGGTGTGGCGC 240
 DB 1601 CTGGTGGGTTTCGCGCAGCAGCATTTGTCAGCCGCTGTCCAGCCCAAGGTGTGGCGC 1660
 QY 241 TGGATCAACGTCGTGTCGGCAGTTGTGATGACCGCATTCGGCATCAAACTGATGTTGATG 300
 DB 1661 TGGATCAACGTCGTGTCGGCAGTTGTGATGACCGCATTCGGCATCAAACTGATGTTGATG 1720
 QY 301 GGTAGTTTTTCGCGGTTTGGATCGGTGCGCTTCGCCCAAAATGTTGATCGCGCGTCG 360
 DB 1721 GGTAGTTTTTCGCGGTTTGGATCGGTGCGCTTCGCCCAAAATGTTGATCGCGCGTCG 1780
 QY 361 TGGGAATCTCATCGATCGCTCCAATTCGCGCTCAGAAAACCTCCAAGTTGTTGAGTGAA 420
 DB 1781 TGGGAATCTCATCGATCGCTCCAATTCGCGCTCAGAAAACCTCCAAGTTGTTGAGTGAA 1840
 QY 421 TCAAGGCTGTTGTCAGCTGCTCAACTGACGAAGCAACCAATCAATGCACTGGTCAACGGTA 480
 DB 1841 TCAAGGCTGTTGTCAGCTGCTCAACTGACGAAGCAACCAATCAATGCACTGGTCAACGGTA 1900
 QY 481 TCCGCGCGTACTCTCTTGTTCGCGCAGCACCATGCAAGCGGCATCTGCGCAAGTGAC 540
 DB 1901 TCCGCGCGTACTCTCTTGTTCGCGCAGCACCATGCAAGCGGCATCTGCGCAAGTGAC 1960
 QY 541 TCCGCGCGTCTCTCTGCGCGATGCTATTGAGCTTCGCGACCATCAATATTGTTCCACGTT 600

DB 1961 TGCCCGGTTCTGGCGGATGTCATTGAGCTTGCGGACCATATCATATATTGTTACGTTTC 2020
 QY 601 AACATGCCCTCAGACAGAGGACTTACCTCGCTGGCGGGGAACCCCTCTTGGAAATTCATCG 660
 DB 2021 AACATGCCCTCAGACAGAGGACTTACCTCGCTGGCGGGGAACCCCTCTTGGAAATTCATCG 2080
 QY 661 AGATATTTGTCGTGAGCAGAGGCTTCGCGCAAGTGTGTGAGAAAGCAATGACGCCAAGACCA 720
 DB 2081 AGATATTTGTCGTGAGCAGAGGCTTCGCGCAAGTGTGTGAGAAAGCAATGACGCCAAGACCA 2140
 QY 721 TTCTTGCGAGCTGACTGCAACCAAGTTCTCACCGTCATCGCCCGGTTCTCCACCCCAACGA 780
 DB 2141 TTCTTGCGAGCTGACTGCAACCAAGTTCTCACCGTCATCCCGCGGTTCTCCACCCCAACGA 2200
 QY 781 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCCGCATGAAC 840
 DB 2201 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCCGCATGAAC 2260
 QY 841 TCAGCCGCTTCGCTGTGAGCTCTGACCGTAG 873
 DB 2261 TCAGCCGCTTCGCTGTGAGCTCTGACCGTAG 2293

RESULT 2

AAH68528/c
 ID AAH68528 standard; DNA; 349980 BP.

XX AC AAH68528;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 7063.

XX DX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN BP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX XX WPI; 2001-376931/40.

XX XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

XX PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX
SQ Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;
Query Match 100.0%; Score 873; DB 22; Length 349980;
Best Local Similarity 100.0%; Pred. No. 6.1e-246;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGCACCTGGTTG 60
DB 128548 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGCACCTGGTTG 128489
QY 61 AACCCGAATGCGATTGAGCGGTTGTTATCGCGCGCGTCTCGCGCGCAATACGGC 120
DB 128488 AACCCGAATGCGATTGAGCGGTTGTTATCGCGCGCGTCTCGCGCGCAATACGGC 128429
QY 121 GACACCGGAGCGGTTTTCGCGCGTCTCGCGCGTCTCGCGCGCAATGTTGTTCCG 180
DB 128428 GACACCGGAGCGGTTTTCGCGCGTCTCGCGCGTCTCGCGCGCAATGTTGTTCCG 128369
QY 181 CTGTGGGTTTTCGCGCGAGCAGCATTTGTACGCGCGTCTCGCGCGCAATGTTGTTCCG 240
DB 128368 CTGTGGGTTTTCGCGCGAGCAGCATTTGTACGCGCGTCTCGCGCGCAATGTTGTTCCG 128309
QY 241 TGGATCAACGTCGTCGTCGAGTTGTGATGACCGCATTTGGCCATCAACTGATTTGATG 300
DB 128308 TGGATCAACGTCGTCGTCGAGTTGTGATGACCGCATTTGGCCATCAACTGATTTGATG 128249
QY 301 GGTAGTTTTCGCGGTTTGGATCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 360
DB 128248 GGTAGTTTTCGCGGTTTGGATCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 128189
QY 361 TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAACTCCAAAGTTGTCAGTGAA 420
DB 128188 TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAACTCCAAAGTTGTCAGTGAA 128129
QY 421 TCAAGGCTGTGTCCAGTCTCAACTGAGGAGCAACCAATCAATGCACTGGTCACGGTA 480
DB 128128 TCAAGGCTGTGTCCAGTCTCAACTGAGGAGCAACCAATCAATGCACTGGTCACGGTA 128069
QY 481 TCGCGCGCTACTCTCTCTGTCGCGCAGCAGCACCATCGAGCGCATCTCGGCAAGTGAC 540
DB 128068 TCGCGCGCTACTCTCTCTGTCGCGCAGCAGCACCATCGAGCGCATCTCGGCAAGTGAC 128009
QY 541 TGCCCGGTTCTCGGCGAGTGTCAATGAGCTTGGGACCATATCAATATTGTTACGTTTC 600
DB 128008 TGCCCGGTTCTCGGCGAGTGTCAATGAGCTTGGGACCATATCAATATTGTTACGTTTC 127949
QY 601 AACATGCCCTCAGACAGGACTTACCTCGTGGCGGGAACCCCTCTGGAATTTCCATCG 660
DB 127948 AACATGCCCTCAGACAGGACTTACCTCGTGGCGGGAACCCCTCTGGAATTTCCATCG 127889
QY 661 AGATATTGTCGTCGAGCAGCCCTCGCGAAGTGTGAGAAAGCAATGACCAAGACCA 720
DB 127888 AGATATTGTCGTCGAGCAGCCCTCGCGAAGTGTGAGAAAGCAATGACCAAGACCA 127829
QY 721 TTGTTGACGCTGACTCAACAAGTTCTCACCGTCATCGCGCGGTCCTCCACCAACGA 780
DB 127828 TTGTTGACGCTGACTCAACAAGTTCTCACCGTCATCGCGCGGTCCTCCACCAACGA 127769
QY 781 TTAATGATGAATAGCTTGGCTGATGAATCAGAGCGGCGAGCCCTCTCCGCGCATGAAC 840
DB 127768 TTAATGATGAATAGCTTGGCTGATGAATCAGAGCGGCGAGCCCTCTCCGCGCATGAAC 127709
QY 841 TCAGCGCCCTCCGCTGAGCTCTGAGACCGTAG 873
DB 127708 TCAGCGCCCTCCGCTGAGCTCTGAGACCGTAG 127676

RESULT 3
AAH68419/c
ID AAH68419 standard; DNA; 627 BP.

XX
AC
XX

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 3454.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX BP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 95JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX P-PSDB; AAG93200.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing
XX expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 1; SEQ ID NO: 3454; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a nucleic acid described
XX in the exemplification of the invention.

XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX SQ Sequence 627 BP; 139 A; 159 C; 185 G; 144 T; 0 other;

Query Match 65.1%; Score 568; DB 22; Length 627;

Best Local Similarity 100.0%; Pred. No. 3.4e-157;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 GTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCCCAATGTTGATGCCGGCTCTGGGA 365
DB 627 GTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCCCAATGTTGATGCCGGCTCTGGGA 568

QY 366 AATCTCATGATCGGCTCCACTCGGCTCAGAAATCTCCAGTTGTTGAGTGAATCAAG 425
DB 567 AATCTCATGATCGGCTCCACTCGGCTCAGAAATCTCCAGTTGTTGAGTGAATCAAG 508

QY 426 GCTGTTGTCCAGCTGCTCAACTGACGAAGACCAATCAATGACCTGGTCACTGGTATCGC 485
DB 507 GCTGTTGTCCAGCTGCTCAACTGACGAAGACCAATCAATGACCTGGTATCGC 448

QY 486 GCCGTACTCTCTTCTCGCGCAGCACCACCATGCAAGCCCATCTGCGCAAGTGAATCGCC 545
DB 447 GCCGTACTCTCTTCTCGCGCAGCACCACCATGCAAGCCCATCTGCGCAAGTGAATCGCC 388

QY 546 GCGTTCCTGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTTGTTACGTTCAACAT 605
 DB |||||
 QY 387 GCGTTCCTGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTTGTTACGTTCAACAT 328
 DB |||||
 QY 606 GCCCTCAGACAGGAGCTTACCTGGCTGGCGGCGGAAACCTCTCTGGAATTCATCGAGATA 665
 DB |||||
 QY 327 GCCCTCAGACAGGAGCTTACCTGGCTGGCGGCGGAAACCTCTCTGGAATTCATCGAGATA 268
 DB |||||
 QY 666 TTTGTCGTGAGCAGGCGCTTGGCGGAGTGTGAGAAAGCAATGACGCGCAAGACATTTT 725
 DB |||||
 QY 267 TTTGTCGTGAGCAGGCGCTTGGCGGAGTGTGAGAAAGCAATGACGCGCAAGACATTTT 208
 DB |||||
 QY 726 GCGAGCTGACTGCAACAAGTTCTCACGTCATCGCCGGTTCTCCACCCCAAGATTAAT 785
 DB |||||
 QY 207 GCGAGCTGACTGCAACAAGTTCTCACGTCATCGCCGGTTCTCCACCCCAAGATTAAT 148
 DB |||||
 QY 786 GATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCAAGACTCAGC 845
 DB |||||
 QY 147 GATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCAAGACTCAGC 88
 DB |||||
 QY 846 GCGCTCGCTGTGAGCTCTGAGACCGTAG 873
 DB |||||
 QY 87 GCGCTCGCTGTGAGCTCTGAGACCGTAG 60
 DB |||||

RESULT 4

AAF68077/c
 ID AAF68077 standard; DNA; 993 BP.
 AC AAF68077;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:669.
 XX
 KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering; db.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN W0200100805-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000MO-IB00926.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031563.
 PR 09-JUL-1999; 99DE-1032122.
 PR 09-JUL-1999; 99DE-1032124.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032182.
 PR 09-JUL-1999; 99DE-1032190.
 PR 09-JUL-1999; 99DE-1032191.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032212.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.

PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 31-AUG-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI; 2001-071486/08.
 DR P-PSDB; AAB76844.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation.
 XX
 PS Claim 3; Page 1107-1108; 1119pp; English.
 XX
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 993 BP; 222 A; 247 C; 277 G; 247 T; 0 other;
 XX
 Query Match 46.3%; Score 404.2; DB 22; Length 993;
 Best Local Similarity 98.8%; Pred. No. 7.9e-109;
 Matches 418; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
 QY 453 AGCACCAATCAATGACTGTGTCACGGTA--TCGGCGCGTACTCTCTTGTCTCGCGCAGC 510
 DB |||||
 QY 993 AGCACCAATCAATGACTGTGTCACGGTAATCCGCGCGCGTACTCTCTTGTCTCGCGCAGC 934
 DB |||||
 QY 511 ACCATGCAAGCGCCATCTGCGCAAGTACTGCCCGGTTCTTGGCGGATGTCATTGAGC 570
 DB |||||
 QY 933 ACCATGCAAGCGCCATCTGCGCAAGTACTGCCCGGTTCTTGGCGGATGTCATTGAGC 874
 DB |||||
 QY 571 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGACTTACCCCTGG 630
 DB |||||
 QY 873 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGACTTACCCCTGG 814
 DB |||||
 QY 631 CTGGCGCGGGAACCCCTCTGGAATTTCCATCGAGATATTTGTCGTGAGCAGGCCCTGCGCA 690
 DB |||||
 QY 813 CTGGCGCGGGAACCCCTCTGGAATTTCCATCGAGATATTTGTCGTGAGCAGGCCCTGCGCA 754
 DB |||||
 QY 691 AGTGGTGAGAAAGCAATGACGCGCAAGACCATTTGTTGGCAGCTGACTGCAACAAGTTCTCA 750
 DB |||||
 QY 753 AGTGGTGAGAAAGCAATGACGCGCAAGACCATTTGTTGGCAGCTGACTGCAACAAGTTCTCA 694
 DB |||||
 QY 751 CGGTCAATCGCCGGTTCTTCCACCCCAAGATTAATGATGAATAGCTTGGCTGATGAATC 810
 DB |||||
 QY 693 CGGTCAATCGCCGGTTCTTCCACCCCAAGATTAATGATGAATAGCTTGGCTGATGAATC 634
 DB |||||
 QY 811 AGAAGCGGCGAGCCCTCTCGGCATGAATCAGCGCCCTCGCTGTGAGCTCTGAGCCG 870
 DB |||||
 QY 633 AGAAGCGGCGAGCCCTCTCGGCATGAATCAGCGCCCTCGCTGTGAGCTCTGAGCCG 574
 DB |||||
 QY 871 TAG 873
 DB |||||

```

Db 573 TAG 571
RESULT 5
AAF71779
ID AAF71779 standard; DNA; 822 BP.
XX
AC AAF71779;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.
XX
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
KW fine chemical production; microorganism; organic acid; nucleoside;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
XX
OS Corynebacterium glutamicum.
XX
XX WO200100843-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX 25-JUN-1999; 99US-0141031.
XX 01-JUL-1999; 99DE-1030476.
XX 02-JUL-1999; 99US-0142101.
XX 08-JUL-1999; 99DE-1031415.
XX 08-JUL-1999; 99DE-1031418.
XX 08-JUL-1999; 99DE-1031419.
XX 08-JUL-1999; 99DE-1031420.
XX 08-JUL-1999; 99DE-1031424.
XX 08-JUL-1999; 99DE-1031428.
XX 08-JUL-1999; 99DE-1031434.
XX 08-JUL-1999; 99DE-1031435.
XX 08-JUL-1999; 99DE-1031443.
XX 08-JUL-1999; 99DE-1031453.
XX 08-JUL-1999; 99DE-1031457.
XX 08-JUL-1999; 99DE-1031465.
XX 08-JUL-1999; 99DE-1031478.
XX 08-JUL-1999; 99DE-1031510.
XX 08-JUL-1999; 99DE-1031541.
XX 08-JUL-1999; 99DE-1031573.
XX 08-JUL-1999; 99DE-1031592.
XX 08-JUL-1999; 99DE-1031632.
XX 08-JUL-1999; 99DE-1031634.
XX 08-JUL-1999; 99DE-1031636.
XX 08-JUL-1999; 99DE-1032125.
XX 08-JUL-1999; 99DE-1032126.
XX 08-JUL-1999; 99DE-1032130.
XX 08-JUL-1999; 99DE-1032186.
XX 08-JUL-1999; 99DE-1032206.
XX 08-JUL-1999; 99DE-1032227.
XX 08-JUL-1999; 99DE-1032228.
XX 08-JUL-1999; 99DE-1032229.
XX 08-JUL-1999; 99DE-1032230.
XX 08-JUL-1999; 99DE-1032922.
XX 08-JUL-1999; 99DE-1032926.
XX 14-JUL-1999; 99DE-1032928.
XX 14-JUL-1999; 99DE-1033004.
XX 14-JUL-1999; 99DE-1033005.
XX 14-JUL-1999; 99DE-1033006.
XX 14-JUL-1999; 99US-0148613.
XX 12-AUG-1999; 99DE-1040764.
XX 27-AUG-1999; 99DE-1040765.
XX 27-AUG-1999; 99DE-1040766.
XX 27-AUG-1999; 99DE-1040832.
XX 31-AUG-1999; 99DE-1041378.
XX 31-AUG-1999; 99DE-1041379.
XX 31-AUG-1999; 99DE-1041380.

PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.
XX
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX WPI; 2001-137957/14.
XX P-PSDB; AAB79660.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
XX pathway proteins, useful for producing fine chemicals in
XX microorganisms, including organic acids, nonproteinogenic amino acids,
XX and purine and pyrimidine bases -
XX
XX Claim 3; Page 233-234; 1737pp; English.
XX
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
XX pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
XX MP nucleic acids are useful for the production of fine chemicals
XX in microorganisms, including organic acids, nonproteinogenic amino
XX acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
XX saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
XX compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;
XX
XX Query Match 37.3%; Score 326; DB 22; Length 822;
XX Best Local Similarity 100.0%; Pred. No. 8.2e-86;
XX Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GTCGATAAGCAGCGGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGTGACCTGGTTG 60
XX Db 497 GTCGATAAGCAGCGGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGTGACCTGGTTG 556
XX QY 61 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGGCGCGTGGCGGCAATACGGC 120
XX Db 557 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGGCGCGTGGCGGCAATACGGC 616
XX QY 121 GACACCGGACGCGTGGATTTCGCGCGTGGCGGCAAGCTGATCTGTTCCCG 180
XX Db 617 GACACCGGACGCGTGGATTTCGCGCGTGGCGGCAAGCTGATCTGTTCCCG 676
XX QY 181 CTGGTGGGTTTCGGCGCAGCAGCATTTGTCAACCGCGCTGTCCAGCCCAAGGTGGCGC 240
XX Db 677 CTGGTGGGTTTCGGCGCAGCAGCATTTGTCAACCGCGCTGTCCAGCCCAAGGTGGCGC 736
XX QY 241 TGGATCAACGTCGTGTGGCAGTTGATGACCGGATGGGCCATCAAACTGATGTTGATG 300
XX Db 737 TGGATCAACGTCGTGTGGCAGTTGATGACCGGATGGGCCATCAAACTGATGTTGATG 796
XX QY 301 GGTTAGTTTTCGCGGGTTTGGGAATC 326
XX Db 797 GGTTAGTTTTCGCGGGTTTGGGAATC 822
XX
XX RESULT 6
XX AAS96098
XX ID AAS96098 standard; DNA; 822 BP.
XX
XX AC AAS96098;
XX
XX 26-FEB-2002 (first entry)
XX

```

XX DE C. glutamicum gene #23 encoding metabolic pathway protein.
 XX DE Metabolic pathway protein; MP; lysine biosynthesis pathway;
 KW methionine biosynthesis pathway; large-scale production of fine chemical;
 KW Corynebacterium diphtheriae; diphtheria; ds.
 XX XX
 OS Corynebacterium glutamicum.
 XX WO200166573-A2.
 XX 13-SEP-2001.
 XX XX
 XX 22-DEC-2000; 2000WO-IB02035.
 XX 09-MAR-2000; 2000US-187970P.
 XX 23-JUN-2000; 2000US-0606740.
 XX XX
 PA (BADI) BASP AG.
 XX Pompejus M, Kroeger B, Schroeder H, Zeider O, Haberhauer G, Kim J;
 PI Lee H, Hwang B;
 XX WPI; 2001-582269/65.
 DR P-PSDB; AAU71888.
 XX XX
 XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
 PT glutamicum, useful for producing methionine and lysine in
 PT Corynebacterium and Brevibacterium -
 XX XX
 PS Disclosure; Page 214-215; 316pp; English.
 XX XX
 CC The present invention relates to the isolation of novel Corynebacterium
 CC glutamicum genes encoding metabolic pathway (MP) proteins
 CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention
 CC include enzymes involved in the lysine and methionine biosynthetic
 CC pathways. The polynucleotide sequences of the invention can be used
 CC for the large-scale production and/or modulation of expressions of the
 CC fine chemicals such as lysine and methionine. The sequences of the
 CC invention may be used to identify C. glutamicum and related organisms
 CC e.g. C. diphtheriae in a subject to detect diphtheria.
 CC AA596073-AA596132 represent C. glutamicum genes encoding the novel
 CC metabolic pathway proteins of the invention.
 XX XX
 SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;
 Query Match 37.3%; Score 326; DB 23; Length 822;
 Best Local Similarity 100.0%; Pred. No. 8.2e-86;
 Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGACCTGTTG 60
 DB 497 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGACCTGTTG 556
 QY 61 AACC CGAATCGTATTGGAGCGGTTTGGTATCGCGCGCTGCGCGCAATACGGC 120
 DB 557 AACC CGAATCGTATTGGAGCGGTTTGGTATCGCGCGCTGCGCGCAATACGGC 616
 QY 121 GACACCGGACGGTGATTTTCGCCGCTGGCGGCTTCGCGCAAGCCTGATCTGTTCCCG 180
 DB 617 GACACCGGACGGTGATTTTCGCCGCTGGCGGCTTCGCGCAAGCCTGATCTGTTCCCG 676
 QY 181 CTGTTGGGTTTCGGCGGACGACATTTGTCAGCGCGCTGTCAGCGCCCAAGGTGTGGCG 240
 DB 677 CTGTTGGGTTTCGGCGGACGACATTTGTCAGCGCGCTGTCAGCGCCCAAGGTGTGGCG 736
 QY 241 TGGATCAACGTCGTCGTCGAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATG 300
 DB 737 TGGATCAACGTCGTCGTCGAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATG 796
 QY 301 GGTAGTTTTCGGCGGTTTTCGGAATC 326
 DB 797 GGTAGTTTTCGGCGGTTTTCGGAATC 822

RESULT 7

AAH68420

ID AAH68420 standard; DNA; 708 BP.

XX XX

AC AAH68420;

XX XX

DT 26-SEP-2001 (first entry)

XX XX

DE C glutamicum coding sequence fragment SEQ ID NO: 3455.

XX XX

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;

XX XX

KW organic acid synthesis; ds.

XX OS

OS Corynebacterium glutamicum.

XX PN

PN EP1108790-A2.

XX PD

PD 20-JUN-2001.

XX PF

PF 18-DEC-2000; 2000EP-0127688.

XX PR

PR 16-DEC-1999; 99JP-0377484.

XX PR

PR 07-APR-2000; 2000JP-0159162.

XX PR

PR 03-AUG-2000; 2000JP-0280988.

XX PA

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX PI

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR

DR WPI; 2001-376931/40.

XX DR

DR P-PSDB; AAG93201.

XX PT

PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying

XX PT

PT mutation point of a gene, measuring expression of a gene, analysing

XX PT

PT expression profile or pattern of a gene and identifying homologous gene

XX PS

PS Claim 1; SEQ ID NO: 3455; 246pp + Sequence Listing; English.

XX XX

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

XX SQ

SQ Sequence 708 BP; 134 A; 173 C; 221 G; 180 T; 0 other;

Query Match 34.7%; Score 303; DB 22; Length 708;

Best Local Similarity 100.0%; Pred. No. 4.6e-79;

Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGACCTGTTG 60

DB 406 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTGACCTGTTG 465

QY 61 AACC CGAATCGTATTGGAGCGGTTTGGTATCGCGCGCTGCGCGCAATACGGC 120

DB 466 AACC CGAATCGTATTGGAGCGGTTTGGTATCGCGCGCTGCGCGCAATACGGC 525

QY 121 GACACCGGACGGTGATTTTCGCCGCTGGCGGCTTCGCGCAAGCCTGATCTGTTCCCG 180

DB 526 GACACCGGACGGTGATTTTCGCCGCTGGCGGCTTCGCGCAAGCCTGATCTGTTCCCG 585

QY 181 CTGTTGGTTTTCGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGGCGC 240
DB 586 CTGTTGGTTTTCGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGGCGC 645
QY 241 TGGATCAACGTCGTCGGCAGTTGTGATGACCCCATTTGGCCATCAAACTGATTTGATG 300
DB 646 TGGATCAACGTCGTCGGCAGTTGTGATGACCCCATTTGGCCATCAAACTGATTTGATG 705
QY 301 GGT 303
DB 706 GGT 708
RESULT 8
AAH45375
ID AAH45375 standard; DNA; 1568 BP.
XX AC AAH45375;
XX AC
DT 11-SEP-2001 (first entry)
XX C. thermoaminogenes lysin biosynthetic enzyme lysE DNA.
XX DE
XX DE
KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;
KW aspartate-semialdehyde dehydrogenase; lysE; ds.
XX OS
XX OS Corynebacterium thermoaminogenes.
XX PN JP2001120270-A.
XX PD
XX PD 08-MAY-2001.
XX PF
XX PF 01-NOV-1999; 99JP-0311148.
XX PR
XX PR 01-NOV-1999; 99JP-0311148.
XX PA (AJIN) AJINOMOTO KK.
XX DR WPI; 2001-364760/38.
XX DR P-PSDB; AAG64047.
XX PT A heat-resistant lysin biosynthetic system enzyme gene of a high
PT temperature-resistant coryneform microbe -
XX PS Example 5; Page 22-24; 27pp; Japanese.
XX CC The invention relates to a gene from a high temperature-resistant
CC coryneform microbe that encodes a heat-resistant lysin biosynthetic
CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity
CC and can be used for growing amino acid-producing microbes. The
CC present sequence encodes an enzyme of the invention.
XX SQ Sequence 1568 BP; 258 A; 525 C; 490 G; 295 T; 0 other;
Query Match 22.2%; Score 193.6; DB 22; Length 1568;
Best Local Similarity 78.4%; Pred. No. 1.2e-46;
Matches 232; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 10 CAGCGGTTTGGTTAAGCCCATTTGATGGCAATCGTGTGACCTGGTTGAACCCGAAT 69
DB 1177 CGGACGTTCTGGTTCAGGCCCATGCTCATGGCCATTTGTGACCTGGCTCAATCCCAT 1236
QY 70 GCCTATTTGGACCGTTTGTGTTTATCGCGCGTGGCGCGCAATACCGGACACCGGA 129
DB 1237 GCCTACTGATGCTCTGCTTCATCGGTGGTGTGGAGCCAGTACGGGGAGACCGGT 1296
QY 130 CGGTGATTTTCGCGTGGCGGTTTCGCGGCAAGCTGATCTGCTTCCGCTGGTGGGT 189
DB 1297 CGGTGATCTTCGCTGGGTGCTTCGCGCCAGCTGTCTGTTCCACTGCTCGGT 1356
QY 190 TTGGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGGCGCTGATCAAC 249

DB 1357 TACGGCGCGCCGACACTGTTCGGCTCCCTGTCTCCAGCCCGGGCTCTGGCGTGGATCAAC 1416
QY 250 GTCGTGTGCGCACTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGTTA 305
DB 1417 ATAGGTGTGCGCGCTGTGTCTCACCGGATTTGGCCGCTGAAGCTGATCTGATGGTTA 1472
RESULT 9
AAQ55755
ID AAQ55755 standard; DNA; 5541 BP.
XX AC AAQ55755;
XX AC
DT 25-SEP-1995 (first entry)
XX DE Escherichia coli genomic probe EC-625.
XX KW Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW K.pneumoniae; E.cloacae; clinical sample; ds.
XX OS Escherichia coli.
XX PN WO9401583-A.
XX PD
XX PD 20-JAN-1994.
XX PF
XX PF 07-JUL-1993; 93WO-JP00936.
XX PR
XX PR 07-JUL-1992; 92JP-0179719.
XX PA (FUSO) FUSO PHARM IND LTD.
XX PA (OHNO) OHNO T.
XX PI Eda S, Matsuhisa A, Ohno T, Uehara H;
XX DR WPI; 1994-035086/04.
XX PT Probe for identifying bacteria causing infectious disease -
PT consists of a DNA fragment obtained by HindIII cleavage of the
PT pathogenic bacterial genomic DNA
XX PS Claim 7; Page 70-73; 100pp; Japanese.
XX CC The nucleotide sequence of a 5541 bp probe obtained by digestion of
CC Escherichia coli genomic DNA with the restriction enzyme HindIII.
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,
CC by HindIII digestion of the genomes of Staphylococcus aureus,
CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
CC to detect their respective microorganisms in clinical samples.
XX SQ Sequence 5541 BP; 1331 A; 1353 C; 1417 G; 1440 T; 0 other;
Query Match 8.2%; Score 71.6; DB 15; Length 5541;
Best Local Similarity 49.4%; Pred. No. 1.8e-10;
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;
QY 203 CATTTGTCAGCCCGCTGTTCACGCCCAAGGTGTGGCTGGATCAACGTCGTCGGCAG 262
DB 2687 CATTATCCATCTGACGCTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGCATGAC 2746
QY 263 TTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGGTAGTTTTCGGGGTTTGG 322
DB* 2747 TCGTCAATCACTGATCTTATGCGGGCGCGGTTCTCGCGCCCTTATTAACAGGTCATT 2806
QY 323 AATCGGTGGCTTTCGCCCAAAATGTTGATGCCGCGCTCGTGGGAAATCTCATCGATCGCT 382
DB 2807 TATCGGAAGACGCTTGCACAGATTCAGTCCCATCGCGATATGCTGATCAATCTGG 2865
QY 383 CCAACTCGGCGTCAAGAACTCAAGTTGTTGATGTAATCAAGGCTGTGTCCAGCTGCT 442
DB 2867 CCAGCTCTCTCGTGCTAAATGTGATTAATTCAGCGCCTGCACGTTCTCTCTCAAGTTGTC 2926

QY 443 CAATGACGAGCACCACCAATCAATGCACTGCTCAGGATATCGGCGCGTACTCTCTTGCT 502
 Db 2927 CGCGCGGC-TGGCCACCAATCAATGCACTGCTCAGGATATCGGCGCGTACTCTCTTGCT 2967
 QY 503 CGCGCAGCACCACCAATGCAAGCGCCATCTGCGCAAGTGAAGTCTGCGCGTCTCTGCGCGATGT 562
 Db 2968 CTTTCAGCAACCAAGCTTAAACGCCATTTGCGCCATTGATTTGTCACGCTGCTGTGCGCATTT 3027
 QY 563 CATTGAGCTTGGCGACCAATATCAATATGTTTCAAGTTCATGCTTCAATGCTTCAAGCGGAGCT 622
 Db 3028 CATTCAATAAGTGTAGGCTGTGAGTGTGGCTTCGCTTAAGCAATTTTCGCGGTCAAGCCAC 3087
 QY 623 TACCCTGGCTGGC-----GCGGAACCTCTGGAATTCATGCAATGATTTGT 670
 Db 3088 GAATTTATTCCTTACGATGCAATCGTGAATTTGCGGAATGCGGTGAGATATTTTC 3147
 QY 671 CCGTGAGCAGCGCTTGGCGAAGTGGTGAAGAAAGCAATGAGCCCAAGCAACCATTTGTCGAG 730
 Db 3148 CGGTGAGCAATCTTGGCGAGAGAGTAAAGCAATACAGCCACGCGGTATTTTGCA 3207
 QY 731 CTGACTGCAACAGTTCTCACCGTATCGCGCGGTCTTCAACCCAAAGCAATTAATGATGG 790
 Db 3208 GGGTATCCAGCA-----GGCGGCTTTTATCCACCCAGCGGTTCAGTAAAT 3252
 QY 791 AATAGCTTGGCTGATGATCAGAGCGGCGAGCTCTCTCGGCATGATCACTACCGCCT 850
 Db 3253 TGTACGAAGGTTGATGAATTAACAGCGGAATTTTCCACTCGCGCAGCACTCAACCAATTT 3312
 QY 851 CCGCTGTGAGCTCTGGACCGTA 872
 Db 3313 TTTGCGTCCGCTCTGGCGAGTA 3334

RESULT 10
 ABK72786/c
 ID ABK72786 standard; DNA; 988 BP.
 AC ABK72786;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Bacillus licheniformis genomic sequence tag (GST) #77.
 XX
 KW Differential gene expression; genomic sequenced tag; GST;
 KW altered culture condition; environmental stress;
 KW physiological provocation; ds.
 XX
 OS Bacillus licheniformis.
 XX
 PN W0200229113-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 05-OCT-2001; 2001WO-US31437.
 XX
 PR 06-OCT-2000; 2000US-0680598.
 PR 27-MAR-2001; 2001US-279526P.
 XX
 PA (NOVO) NOVOZYMES BIOTECH INC.
 PA (NOVO) NOVOZYMES AS.
 XX
 PI Berka R, Clausen IG;
 XX
 DR WPI; 2002-416684/44.
 XX
 PT Monitoring differential expression of several genes in first Bacillus
 PT cell relative to expression of same genes in one or more second
 PT Bacillus cells, by using substrate containing Bacillus genomic
 PT sequenced tag array -
 XX
 PS Claim 4; SEQ ID NO 77; 200pp; English.
 XX
 CC The invention describes a method of monitoring differential expression of

CC genes in a first Bacillus cell relative to expression of the genes in
 CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
 CC isolated from Bacillus cells to a substrate containing array of Bacillus
 CC genomic sequenced tags (GST), examining the array, and determining
 CC relative gene expression by an observed hybridisation reporter signal of
 CC a spot in the array. The method is useful for measuring the expression of
 CC genes in a first Bacillus cell relative to expression of the same genes
 CC in one or more second Bacillus cells. The method is useful for monitoring
 CC global expression of several genes from a Bacillus cell, discovering new
 CC genes, identifying possible functions of unknown open reading frames and
 CC monitoring gene copy number variation and stability. Monitoring changes
 CC in expression of genes may be used to provide a representation of the way
 CC in which Bacillus cells adapt to changes in culture conditions.
 CC environmental stress or other physiological provocation. Extensive
 CC follow-up characterisation is unnecessary, when one spot on an array
 CC equals one gene or one open reading frame, since sequence information is
 CC available. This sequence represents a genomic sequence tag (GST) used in
 CC the method of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at
 CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 988 BP; 274 A; 247 C; 254 G; 213 T; 0 other;

Query Match 7.5%; Score 65.4; DB 24; Length 988;
 Best Local Similarity 55.3%; Pred. No. 5.9e-09;
 Matches 204; Conservative 0; Mismatches 141; Indels 24; Gaps 3;

QY 511 ACCCATGCAAGCGCCATCTGCGCAAGTGAAGTCTGCGCGTCTGCGCGATGATTCAGC 570
 Db 857 AGCCAGGCTATCGCCATTTGCGGAAGAGACTGCGCGGTTCGCGACCCACTTGATTCAGC 798
 QY 571 TTGCGGACCATATCAATATTTGTTCAAGTTCACGTTCAACATGCCCTCAGACAGGACTTAC----- 625
 Db 797 TTGCGGACCTTGGCGAGCAGCTTCTGTGATGCCCTCGGCCCTGCAAAACGAGCTGTTTC 738
 QY 626 -CTGCGTGGCGCGGAACCTCTGGAATTCATCGAGATATTTGTCGCTGAGCAGGCC 684
 Db 737 GCATCAGCGCTCTTGACCGCGGGAATCCCTGTCAAAATATTTCTGTTTAAAGCCCT 678
 QY 685 TGGCAAGTGGTGAGAAAGCAATGACGCCAAGACCAATGTTGCGAGCTGACTGCAACAAG 744
 Db 677 TGGCAAGAGGTGTAAACCGGATCGCCCGGCTCTTTTT-----CTGACGACT 627
 QY 745 TTCTCACCGTCACTGCGCGTCTCTCCACCCAAAGATTAATGATGAATAGCTTGCTGA 804
 Db 626 TGCAGCAGCGCTTTTTCGCGTCT-----CTGTGAACATTTGAATATTTGGGCTGA 576
 QY 805 TGAATCAGAGCGGCGAGCGCTCTCTCGGCATGAATCTAGCGCCCTCCCGCTGAGCTCT 864
 Db 575 TGGATGAGGAGCGGAATCCCGTATTCTTGCMAAATATCAGCGCCCTTTTCGGTCTGTTCA 516
 QY 865 GGACCGTAG 873
 Db 515 GCACTGTAG 507

RESULT 11
 AAS59552/c
 ID AAS59552 standard; DNA; 23128 BP.

XX AAS59552;
 XX

DT 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein encoding DNA #47.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant; ds.

OS Propionibacterium acnes.
 PN W0200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US12865.
 XX
 PR 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 DR WPI; 2001-616774/71.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX
 PS Claim 1; SEQ ID No 47; 1069pp; English.
 XX
 CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
 CC associated DNA sequences are used in the treatment, prevention and
 CC diagnosis of medical conditions caused by P. acnes. The disorders include
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved
 CC in infections of bone, joints and the central nervous system, however it
 CC is particularly involved in the inflammatory lesions associated with acne
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a
 CC patient comprises contacting a sample with a binding agent that binds to
 CC the proteins of the invention and determining the amount of bound protein
 CC in the sample. The polypeptides may be used as antigens in the production
 CC of antibodies specific for P. acnes proteins. These antibodies can be
 CC used to downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
 CC polypeptides shown in AAU51663-AAU51893 and AAU67535.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 23128 BP; 4349 A; 6746 C; 7113 G; 4908 T; 12 other;
 XX
 Query Match 6.6%; Score 57.6; DB 23; Length 23128;
 Best Local Similarity 51.6%; Pred. No. 4.5e-06;
 Matches 194; Conservative 1; Mismatches 160; Indels 21; Gaps 2;
 XX
 QY 500 GCTCGGAGACCCATGCAAGCGCCATCTGCGCAAGTACTGCCCGTCTCGGCGGA 559
 Db 7924 GCTCAGCAGGATGACGAGAGGCGCATCTGGGCAACGCTGTGTCACGGTCTTSGCAA 7865
 QY 560 TGTCATTGAGCTTCGGGACCATATCATATTTGTCAGTTCAACATGCCCTCAGACAGG 619
 Db 7864 CCTCGTTGAGGCGAGCAATTTGACGCGGGGTGACGTGCTGATCTGATCGGTCGACGG 7805
 QY 620 A---CTTACCTGCTGGCGCGGAACCCCTCTGGAATTTCCATCAGATATTGTTCGCTGA 676
 Db 7804 AGACCGCTCCCATGCCAGACGCGACTGGGGGGGAACACCCCTCAGTATTGTTCGGTGA 7745
 QY 677 GCAGGCCCTCGGCAAGTGGTGAAGAACAATGACGCCAAGACCACTTTGTCAGCTGACT 736
 Db 7744 GGATTCTCGTGAAGCAACGAGAGAAACAGATGATGCCATCCCTCTCTGCTCAGGTAT 7685
 QY 737 GCACAAAGTTCTACCGTCAATCGCCCGGTCTCTCCACCCCAACGATTAATGATGAATAGC 796
 Db 7684 CGA-----CGAGACCATCTCTCAATCCATCGATCAACATGTTGTAAC 7643

QY 797 TTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCAACTCAGCGCGCTCGCTG 856
 Db 7642 GCGGCTGGTGGATGAGTACGGGGTCCCAAGTACGCGCAATGGTGGCAGACGTTCTG 7583
 QY 857 TGAGCTCTGGACCGTA 872
 Db 7582 TCTCCTTGGGATCGTA 7567
 RESULT 12
 AAT33536/c
 ID AAT33536 standard; DNA; 15239 BP.
 XX
 AC AAT33536;
 XX
 DT 15-FEB-1998 (first entry)
 XX
 DE BCG deletion region 2 and flanking sequences.
 XX
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;
 KW mycobacteria; vaccine; infection; marker; ss.
 XX
 OS Mycobacterium bovis strain BCG.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 3382..14071
 FT /*tag= a
 FT /*note= "BCG delta 1 deletion region"
 XX
 PN W09625519-A1.
 XX
 PD 22-AUG-1996.
 XX
 PF 15-FEB-1996; 96WO-US01938.
 XX
 PR 17-FEB-1995; 95US-0390878.
 XX
 PA (PATH-) PATHOGENESIS CORP.
 XX
 PI Mahairas GG, Stover CK;
 XX
 DR WPI; 1996-393419/39.
 XX
 PT Detecting markers for avirulence in Mycobacterium - used in
 PT production of vaccines against bacterial infection, and to detect
 PT bacterial infection
 XX
 PS Example 1; Fig 2; 66pp; English.
 XX
 CC This DNA sequence comprises Mycobacterium bovis BCG deletion
 CC sequence BCGdelta2. A specific genetic deletion of this region
 CC results in an avirulence phenotype of the mycobacterium. 2 Other
 CC deletion regions (see AAT33535 and AAT33537) have also been detected.
 CC Identification involved screening a BCG cosmid library with a
 CC radiolabeled probe obtained following DNA subtraction between
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.
 CC The deletions provide useful markers for the identification of an
 CC avirulent, or a virulent, mycobacterial phenotype. Determination
 CC of avirulence requires the detection of the presence or absence of
 CC the deletion; the deletions are detected either by detecting the
 CC presence or absence of deletion junctions (see AAT33538-46), or by
 CC detecting the presence or absence of the sequences contained within
 CC the deletion. Deletion polypeptides are used as components of
 CC immunological assays and in vaccines.
 XX
 SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;
 XX
 Query Match 6.5%; Score 57; DB 17; Length 15239;
 Best Local Similarity 53.1%; Pred. No. 5.6e-06;
 Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;
 QY 26 AGCCCATGTTGATGGCAATCGTGTGACCTGTTGAACCCGAATCGTATTGGACGCT 85

Db 4792 AGGTCCTGGTGACCTGTGGCGCATTCACGTTCTCTCAACCCACACAGCTTACTCTGCACACCG 4733
 QY 86 TTGTTGTTTATCGCGCGGTGGCGCGCAATACGGCGACACGGACGCTGATTTTCGCCG 145
 Db 4732 TCGTGTGCTAGGCGCGCTGGCCAAACAGACACAGCGA---CCAGCGCTGGCTGTTCGCCG 4676
 QY 146 CTGCGCGGTTTCGCGCGCAAGCTGATCTGTTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT 205
 Db 4675 TCGCGCGGTTTCACAGCAGTGGGTATGTTTCGCCACCTCGGGTTTCGGAGCCGCGGT 4616
 QY 206 TGTCAACCCCGCTGTTCAGACCCCAAGGTGTGGCGCTGGATCAACGTCGTGGCGCATTTG 265
 Db 4615 TCGCGCGGCTGTTTCAACCAACCCCGCTGTGGAGAACTCTCGAGCGCTGATCGCGGTCA 4556
 QY 266 TGATGACCGCATTTGGCCATCAAACTGATTTGA 298
 Db 4555 TGATGGTTGGCTGGGAATCTCGCTGACCCGTGA 4523

RESULT 13

AAI99683
 ID AAI99683 standard; DNA; 4403765 BP.

XX AC AAI99683;

DT 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 PT 1551 and H37Rv differ

XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
 Query Match 6.5%; Score 57; DB 22; Length 4403765;
 Best Local Similarity 53.1%; Pred. No. 6.3e-05;

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;
 QY 26 AGCCCATGTTGATGGCAATCGTCTGCTGACCTGTTGAACCCGAATGCTATTTGGACGGT 85
 Db 2227633 AGGTCTCTGTGACCTGTGGCGCATTCACGTTCTCAACCCACACGCTCTACCTCGACACCG 2227692
 QY 86 TTGTTGTTTATCGCGCGGTGGCGCGCAATACGGCGACACACCGAGCTGTGATTTTCGCCG 145
 Db 2227693 TCGTGTGCTAGGCGCGCTGGCCAAACAGACACAGGA---CCAGCGCTGGCTGTTCGCCG 2227749
 QY 146 CTGCGCGGTTTCGCGCGCAAGCTGATCTGTTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT 205
 Db 2227750 TCGCGCGGTTTCACAGCAGTGGGTATGTTTCGCCACCTCGGGTTTCGGAGCCGCGGT 2227809
 QY 206 TGTCAACCCCGCTGTTCAGACCCCAAGGTGTGGCGCTGGATCAACGTCGTGGCGAGTTG 265
 Db 2227810 TCGCGCGGCTGTTCAACCAACCCCGCTCGTGGAGAACTCTCGACGGCTGATCGCGGTCA 2227869
 QY 266 TGATGACCGCATTTGGCCATCAAACTGATTTGA 298
 Db 2227870 TGATGGTTGGCTGGGAATCTCGCTGACCCGTGA 2227902

RESULT 14

AAI99682
 ID AAI99682 standard; DNA; 4411529 BP.

XX AC AAI99682;

DT 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 PT 1551 and H37Rv differ

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX

SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;
Query Match 6.5%; Score 57; DB 22; Length 4411529;
Best Local Similarity 53.1%; Pred. No. 6.3e-05;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;
QY 26 AGCCCATGTTGATGGCAATCGTGTGACCTGTTGAACCCGAATGCGTATTGGACGCGT 85
DB 2230334 AGGTCTCTGTGACCTGTGCGGCATTACGTTCTCAACCCACACAGCTTACCTCGACACCG 2230393.
QY 86 TTGTTGTTATCGCGCGGTGCGCGGCAATACGCGGACACCGGACGCGTGTGATTTCGCG 145
DB 2230394 TCGTGTGCTAGGCGCGCTGCGCAACGAGCACAGCGA---CCAGCGCTGCGCTGTCGCGC 2230450
QY 146 CTGCGCGGTTTCGCGCAAGCTGATCTGTTTCCCGTGGTGGTTCGCGCGCAGCAGCAT 205
DB 2230451 TCGCGCGGTTCACAGCAGTGGGTATGTTTCCACCTCGGTTTCGAGCGCGCGGT 2230510
QY 206 TGTCAAGCGCGCTGTCAAGCCCAAGGTGCGCTGATCAACGTGCTGTGCGCAGTTG 265
DB 2230511 TCGCGCGGTGTTCAACCAACCGCGTCTGTGAGAAATCCTCGACGCGCTGATCGCGTCA 2230570
QY 266 TGATGACGCGATTGCGCCATCAAACTGATGTGA 298
DB 2230571 TGAATGCTGCGTGGGAATCTCGTACCGGTGA 2230603

RESULT 15
AAA52691
ID AAA52691 standard; DNA; 636 BP.
AC AAA52691;
XX
XX 03-JAN-2001 (first entry)
XX Escherichia coli yggA gene.
XX
XX E. coli; yggA gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX Escherichia coli.
XX
XX Key Location/Qualifiers
FT CDS 1..636
FT /*tag= a
FT /product= "yggA"
XX
XX EP1016710-A2.
XX
XX 05-JUL-2000.
XX
XX 17-DEC-1999; 99BP-0125263.
XX
XX 30-DEC-1998; 98RU-0124016.
XX 09-MAR-1999; 99RU-0104431.
XX
XX (AJIN) AJINOMOTO CO INC.
XX
XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
XX
XX WPI; 2000-414802/36.
DR P-PSDB; AAB01789.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX
XX Disclosure; Page 24; 29pp; English.
XX
XX The present sequence is the yggA gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its

CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in arginine, glutamic acid and lysine is
CC achieved if multiple copies of the gene are transfected into a bacterium.
CC The bacterium used is E. coli.

XX SQ Sequence 636 BP; 120 A; 143 C; 180 G; 193 T; 0 other;

Query Match 6.5%; Score 56.4; DB 21; Length 636;
Best Local Similarity 52.1%; Pred. No. 2.2e-06;
Matches 126; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 50 TGACCTGGTTGAACCCGAATGCGTATTGGACGCGTGTGTTTATCGCGCGCGTGGCG 109
DB 353 TGACCTGGCTGAATCGCATGTTTACCTGATACCTTTTGTGTACTGGCAGCCTTGGCG 412
QY 110 CGCAATACGGCGACACCGGACGGTGAATTTTCGCCCTCGCGGTTTCGCGGCAAGCCTGA 169
DB 413 GGCAACTTGTATGTGGAACCAAAACGCTGGTTTGCACCTCGGACAAATTAGCGCCTCTTTCC 472
QY 170 TCTGGTTCCCGCTGGTGGTTCGCGCGCAGCAGCATTTGTACGCCGCTGTCAGCCCA 229
DB 473 TGTGGTTCCTTTGGTCTGGCTCTTCTCGCAGCCTGGCTGGCACCGCTCTGGCAGCGCAA 532
QY 230 AGGTGTGGCGCTGGATCAACGTCGTGTCGGCAGTTGTGTATGATGACCGCATTTGGCCATCAAAAC 289
DB 533 AAGCACAGCGCATTTATCAATCTGGTTGTGGATGTGTTATGTGTTTATTGCTTGCAGC 592
QY* 290 TG 291
DB 593 TG 594

Search completed: April 27, 2003, 00:06:32
Job time : 5330.11 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 98.593 Seconds
(without alignments)
9634.930 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1421_2293

Perfect score: 873

Sequence: 1 gtcgataagcagcggttg.....ctgtgagctgacgcgtag 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT NEW PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	873	100.0	3309400	9	US-09-738-626-1
C 2	568	65.1	627	9	US-09-738-626-3454
C 3	326	37.3	822	9	US-09-746-660A-51
C 4	303	34.7	708	9	US-09-738-626-3455
C 5	65.4	7.5	988	10	US-09-974-300-77
C 6	57	6.5	597	9	US-09-894-844-11
C 7	50.2	5.8	1277	9	US-09-971-536-15
C 8	40	4.6	399	10	US-09-861-893-29
C 9	40	4.6	431	10	US-09-954-456-1034
C 10	38.2	4.4	948	9	US-09-738-626-1672
C 11	36.4	4.2	499	9	US-10-184-644-592
C 12	36.4	4.2	499	9	US-10-184-634-592
C 13	36	4.1	1068	9	US-09-738-626-439
C 14	36	4.1	3309400	9	US-09-738-626-1
C 15	35.4	4.1	996	10	US-09-815-242-7775
C 16	34.8	4.0	542	9	US-10-184-644-398
C 17	34.8	4.0	542	9	US-10-123-155-188
C 18	34.8	4.0	542	9	US-10-184-634-398
C 19	34.8	4.0	3017	10	US-09-818-143-8

20	34.8	4.0	7158	10	US-09-974-300-2171	Sequence 2171, Ap
21	33.8	3.9	887	10	US-09-864-761-21484	Sequence 21484, A
22	33.8	3.9	1981	10	US-09-864-761-4745	Sequence 4745, Ap
23	33.8	3.9	2930	10	US-09-745-763-198	Sequence 198, App
24	33.6	3.8	1075	10	US-09-864-761-19241	Sequence 19241, A
25	33.6	3.8	1403	10	US-09-864-761-2513	Sequence 2513, Ap
26	33.6	3.8	1446	9	US-09-712-363-57	Sequence 57, Appl
27	33.2	3.8	1635	10	US-09-864-761-20241	Sequence 20241, A
28	33.2	3.8	1973	10	US-09-864-761-3471	Sequence 3471, Ap
29	33	3.8	9210	9	US-09-712-363-100	Sequence 100, App
30	32.8	3.8	2048	10	US-09-925-300-760	Sequence 760, App
31	32.4	3.7	4689	9	US-09-860-846-34	Sequence 34, Appl
32	32.4	3.7	4689	9	US-09-988-3848-34	Sequence 34, Appl
33	32.4	3.7	4689	10	US-09-861-289-34	Sequence 34, Appl
34	32.4	3.7	36778	9	US-09-860-846-5	Sequence 5, Appli
35	32.4	3.7	36778	10	US-09-861-289-5	Sequence 5, Appli
36	32.4	3.7	37948	9	US-09-988-3848-5	Sequence 5, Appli
37	32.2	3.7	484	9	US-09-918-995-3113	Sequence 3113, Ap
38	32.2	3.7	497	9	US-09-918-995-27219	Sequence 27219, A
39	32.2	3.7	653	9	US-10-184-644-402	Sequence 402, App
40	32.2	3.7	653	9	US-10-184-634-402	Sequence 402, App
41	32.2	3.7	1404	9	US-09-976-740-10	Sequence 10, Appl
42	32.2	3.7	1404	10	US-09-962-055-10	Sequence 10, Appl
43	32.2	3.7	1404	12	US-10-023-529-10	Sequence 10, Appl
44	32.2	3.7	1404	12	US-10-023-523-10	Sequence 10, Appl
45	32	3.7	360	9	US-10-184-644-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication NO. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 100.0%; Score 873; DB 9; Length 3309400;
Best Local Similarity 100.0%; Pred. No. 7.6e-281;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGTTTGGTAAAGCCATGTTGATGGCAATCGTGTACCTGGTTG 60
Db_1328548 GTCGATAAGCAGCGGTTTGGTAAAGCCATGTTGATGGCAATCGTGTACCTGGTTG 1328489
QY 61 AACCCGAATGCGTATTTCGACCGGTTTGTGTTATCGCGCGCGCTCGCGCGCAATACGCG 120

Db 1328488 AACCCGATCGTATTTGGAGCGGTTTGTGTTATTCGGCGCGCTCGCGCGCATACGCG 1328429
Qy 121 GACACCGACGGTGGATTTTCGCGCTGGCGGCTTTCGGCGGCAAGCTGATCTGTTCCCG 180
Db 1328428 GACACCGACGGTGGATTTTCGCGCTGGCGGCTTTCGGCGGCAAGCTGATCTGTTCCCG 1328369
Qy 181 CTGTGGGTTTCGCGGACGAGCATTTGTCAGCGCGCTGTCAGCGCCCAAGGTTGCGGC 240
Db 1328368 CTGTGGGTTTCGCGGACGAGCATTTGTCAGCGCGCTGTCAGCGCCCAAGGTTGCGGC 1328309
Qy 241 TGGATCAACGTCGTCGTCGCGAGTTGTGATGACCGCATTTGGCCCAAAATGTTGATCGCGCGCTG 300
Db 1328308 TGGATCAACGTCGTCGTCGCGAGTTGTGATGACCGCATTTGGCCCAAAATGTTGATG 1328249
Qy 301 GGTAGTTTCGCGGCTTTCGGAATCGGTGGCTTTCGCGGCAAAATGTTGATCGCGCGCTG 360
Db 1328248 GGTAGTTTCGCGGCTTTCGGAATCGGTGGCTTTCGCGGCAAAATGTTGATCGCGCGCTG 1328189
Qy 361 TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAAACTCCAAAGTTGTTGAGTGA 420
Db 1328188 TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAAACTCCAAAGTTGTTGAGTGA 1328129
Qy 421 TCAAGGCTGTTGTCAGCTGCTCAACTGACGAAGCACCATCAATGCACTGCTCACGGTA 480
Db 1328128 TCAAGGCTGTTGTCAGCTGCTCAACTGACGAAGCACCATCAATGCACTGCTCACGGTA 1328069
Qy 481 TCGCGCGCTACTCTCTCTGTCGCGAGCACCATGCAAGCGCATCTGCGCAAGTGAC 540
Db 1328068 TCGCGCGCTACTCTCTCTGTCGCGAGCACCATGCAAGCGCATCTGCGCAAGTGAC 1328009
Qy 541 TGCCGCGCTTCTCGCGGATGCTTCAATGAGCTTGGGACCATATCAATATTTGTTCACTTC 600
Db 1328008 TGCCGCGCTTCTCGCGGATGCTTCAATGAGCTTGGGACCATATCAATATTTGTTCACTTC 1327949
Qy 601 AACATGCTTTCAGACAGGAGCTTACCTTGGCTGGCGGGAACCTCTCGGAATTTCCATCG 660
Db 1327948 AACATGCTTTCAGACAGGAGCTTACCTTGGCTGGCGGGAACCTCTCGGAATTTCCATCG 1327889
Qy 661 AGATATTTGTCGTCGAGAGCGCTTTCGCGAGCTTGGGACCATATCAATATTTGTTCACTTC 720
Db 1327888 AGATATTTGTCGTCGAGAGCGCTTTCGCGAGCTTGGGACCATATCAATATTTGTTCACTTC 1327829
Qy 721 TTGTTGGCAGCTGATGCAACAAGTTCTACCGCTCATCGCCCGGTTCTCCACCCCAACGA 780
Db 1327828 TTGTTGGCAGCTGATGCAACAAGTTCTACCGCTCATCGCCCGGTTCTCCACCCCAACGA 1327769
Qy 781 TTAATGATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCATGAAC 840
Db 1327768 TTAATGATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCATGAAC 1327709
Qy 841 TCAGCGGCTTCGCTGTGAGCTCTGGACCGTAG 873
Db 1327708 TCAGCGGCTTCGCTGTGAGCTCTGGACCGTAG 1327676

RESULT 2

US-09-738-626-3454/c
; Sequence 3454, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 3454
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-3454

Query Match 65.1%; Score 568; DB 9; Length 627;

Best Local Similarity 100.0%; Pred. No. 3e-180;

Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 306 GTTTTCGCGGCTTTCGGAATCGGTGGCTTTCGCGGCAAAATGTTGATCGCGCGCTCGTGGGA 365
Db 627 GTTTTCGCGGCTTTCGGAATCGGTGGCTTTCGCGGCAAAATGTTGATCGCGCGCTCGTGGGA 568
Qy 366 AATCTCATGATCGCTCCAACTCGCGGTCAGAAACTCCAAAGTTGTTGAGTGAATCAAG 425
Db 567 AATCTCATGATCGCTCCAACTCGCGGTCAGAAACTCCAAAGTTGTTGAGTGAATCAAG 508
Qy 426 GCTGTTGTCAGCTGCTCAACTGACGAAGCACCATCAATGCACTGCTCACGGTATCCGC 485
Db 507 GCTGTTGTCAGCTGCTCAACTGACGAAGCACCATCAATGCACTGCTCACGGTATCCGC 448
Qy 486 GCGTACTCTCTTGTTCGCGAGCACCATGCAAGCGCATCTGCGCAAGTGAATCGGCC 545
Db 447 GCGTACTCTCTTGTTCGCGAGCACCATGCAAGCGCATCTGCGCAAGTGAATCGGCC 388
Qy 546 GCGTCTTCTGGCGGATGCTCAATGAGCTTTCGCGGCAACCTCTCGGAATTTCCATCGATA 605
Db 387 GCGTCTTCTGGCGGATGCTCAATGAGCTTTCGCGGCAACCTCTCGGAATTTCCATCGATA 328
Qy 606 GCCTTCAGACAGGAGCTTACCTTGGCTGGCGGGAACCTCTCGGAATTTCCATCGATA 665
Db 327 GCCTTCAGACAGGAGCTTACCTTGGCTGGCGGGAACCTCTCGGAATTTCCATCGATA 268
Qy 666 TTTGTCGTCGAGCAGCGCTTCGCAAGTGTGAGAAAGCAATGACGCCAAGACCATGTT 725
Db 267 TTTGTCGTCGAGCAGCGCTTCGCAAGTGTGAGAAAGCAATGACGCCAAGACCATGTT 208
Qy 726 GGCAGCTGATGCAACAAGTTCTCACCGCTCATCGCCGGTTCCTCCACCAAGCATTAAT 785
Db 207 GGCAGCTGATGCAACAAGTTCTCACCGCTCATCGCCGGTTCCTCCACCAAGCATTAAT 148
Qy * 786 GATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCATGAATCAGC 845
Db 147 GATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCGCCCATGAATCAGC 88
Qy 846 CGCTTCGCTGTGAGCTCTGGACCGTAG 873
Db 87 CGCTTCGCTGTGAGCTCTGGACCGTAG 60

RESULT 3

US-09-746-660A-51
; Sequence 51, Application US/0974660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:

; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick

APPLICANT: Hwang, Byung-Joon
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
FILE REFERENCE: BGI-121C2
CURRENT APPLICATION NUMBER: US/09/746,660A
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 09/606740
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 09/603124
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/141031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: 60/142101
PRIOR FILING DATE: 1999-07-02
PRIOR APPLICATION NUMBER: 60/148613
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/187970
PRIOR FILING DATE: 2000-03-09
PRIOR APPLICATION NUMBER: DE 19931420.9
PRIOR FILING DATE: 1999-07-08
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PatentIn Vers. 2.0
SEQ ID NO 51
LENGTH: 822
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(799)
OTHER INFORMATION: RXA01394
US-09-746-660A-51

Query Match 37.3%; Score 326; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 5.4e-99;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 60
Db 497 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 556
QY 61 AACCCGAATCGTATTTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGCG 120
Db 557 AACCCGAATCGTATTTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGCG 616
QY 121 GACACCGAGCGTGGATTTTCGCGCGTGGCGCGTTCGCGGCAAGCTGATCTGGTTCCCG 180
Db 617 GACACCGAGCGTGGATTTTCGCGCGTGGCGCGTTCGCGGCAAGCTGATCTGGTTCCCG 676
QY 181 CTGCTGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCCAGCCCAAGGTGTGGCG 240
Db 677 CTGCTGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCCAGCCCAAGGTGTGGCG 736
QY 241 TGGATCAACGTCGTCGCGCGAGTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATG 300
Db 737 TGGATCAACGTCGTCGCGCGAGTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATG 796
QY 301 GGTTAGTTTCGCGCGGTTTGGGAATC 326
Db 797 GGTTAGTTTCGCGCGGTTTGGGAATC 822

RESULT 4
US-09-738-626-3455
Sequence 3455, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 3455
LENGTH: 708
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3455

Query Match 34.7%; Score 303; DB 9; Length 708;
Best Local Similarity 100.0%; Pred. No. 2.6e-91;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 60
Db 406 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 465
QY 61 AACCCGAATCGTATTTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGCG 120
Db 466 AACCCGAATCGTATTTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGCG 525
QY 121 GACACCGAGCGTGGATTTTCGCGCGTGGCGCGTTCGCGGCAAGCTGATCTGGTTCCCG 180
Db 526 GACACCGAGCGTGGATTTTCGCGCGTGGCGCGTTCGCGGCAAGCTGATCTGGTTCCCG 585
QY 181 CTGCTGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCCAGCCCAAGGTGTGGCG 240
Db 586 CTGCTGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCCAGCCCAAGGTGTGGCG 645
QY 241 TGGATCAACGTCGTCGTCGCGAGTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATG 300
Db 646 TGGATCAACGTCGTCGTCGCGAGTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATG 705
QY 301 GGT 303
Db 706 GGT 708

RESULT 5
US-09-974-300-77/c
Sequence 77, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 77
LENGTH: 988
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-77

Query Match 7.5%; Score 65.4; DB 10; Length 988;
Best Local Similarity 55.3%; Pred. No. 1.6e-11;
Matches 204; Conservative 0; Mismatches 141; Indels 24; Gaps 3;

QY 511 ACCCATGCAAGCGGCATCTGGCGAAGTGACTGCGCGGTTCTCTGGGCGATGTCAATTGAGC 570
DB 857 AGCAGGCTATCGCATTTGGGGAAGAGACTGGCGCGTTGGGAGCAGCATTTGATTGAGC 798
QY 571 TTGGGACCATATCAATATTGTTTCAACATGCGCTTCAACATGCGCTTCAACATGCGCTTCAAC 625
DB 797 TTGGGACCTTGGCGAGCAGCTCTTCTGTGATGCCCTCGCGTCAAAAGCGACTGTTTC 738
QY 626 -CCTGGCTGGGGGGGAACCTCTGGGAATTCATCGAGATATTGTCCTGGTGGAGCGCC 684
DB 737 GCATCAGCGGCTCTTGACCGCGGGAATCCCTGTCAATATTCTGTTGTTAAAGCCCT 678
QY 685 TGCGCAAGTGTGAGAAAGCAATGACGCCAAGCAATGTTGGCAGCTGACTGCAACAAG 744
DB 677 TGGCAAGAGGTGTAAGCGATGCGCGGCTCTTTT-----CTGCGAGCT 627
QY 745 TTCTCAGCGTATCGCGCGGTTCTTCCACCAACGATTAATGATGAATAGTTGGCTGA 804
DB 626 TGCAGCAGGCTTTTTCGGGTTCT-----CTGTGGAACATTGAATATTGGGCTGA 576
QY 805 TGATCAGAGCGGCGAGCGCTCTCGCCATGAACTCAGCGCGCTCGCTGTGAGCTCT 864
DB 575 TGGATGAGGAGCGGAATCCCGTATTCTTGCAAAATATCAGCGCGCTTTTCGGTCTGTTCA 516
QY 865 GGACCGTAG 873
DB 515 GCACTGTAG 507

RESULT 6
US-09-894-844-11
; Sequence 11, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; TITLE OF INVENTION: the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-09-894-844-11

Query Match 6.5%; Score 57; DB 9; Length 597;
Best Local Similarity 53.1%; Pred. No. 8.4e-09;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCCATGTTGATGGCAATGCTGTGACCTGGTTGAACCCGGAATCGGTATTGGACCGCT 85
DB 326 AGGTCTGTTGACCTGTGCGCAATTCAGTTCTTCAACCCACACGTCTACTCTGACACCG 385
QY 86 TTGTGTTTATCGGCGGCTGCGCGGCAATACGCGGACACCGGACGTTGATTTTCGCG 145
DB 386 TCGTGTGTTGCTAGGCGCGCTGCGCAACGAGCACAGCGA---CCAGCGTGGCTGTTGCGCC 442
QY 146 CTGCGCGCTTTCGCGCAAGCGCTGATCTGTTTCCCGCTGGTGGTTCGCGCGCAGCAT 205

DB 443 TGGCGCGGTACAGCCAGTGTGGTATGTTGCGCACCCCTCGGGTTCGAGCGCGCGGT 502
QY 206 TGTACAGCCCGCTGTTCAGCCCCCAAGGTGTGGCTTGGATCAACGTCGTCGTGGCAGTTG 265
DB 503 TGGCGGGCTGTTTACCAACCCCGGCTCGTGAGAATCTCTGACGCGCTGTATCGGGTCA 562
QY 266 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 298
DB 563 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 595

RESULT 7
US-09-971-536-15/c
; Sequence 15, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-15

Query Match 5.8%; Score 50.2; DB 9; Length 1277;
Best Local Similarity 49.3%; Pred. No. 2.3e-06;
Matches 176; Conservative 0; Mismatches 163; Indels 18; Gaps 1;

QY 365 AATCTCATGATCGCCTCCAACCTCGGCTCAGAAAATCCAAAGTTGTTGAGTGAATCAA 424
DB 1156 AATATCATCAATTTGTTGAATCTCTTACGAGTAAAGGTGAGTGTTCGTTGCTTTAA 1097
QY 425 GGCTGTTGTCAGCTCTCAATGACGAGCAGCACCATTCAATGACCTGCTCAGGTATCCG 484
DB 1096 GGTATCTGAGGTGTTCACTGAGGTCTCCCAATGATCACTTTGTGACAA----- 1043
QY 485 CGCCGTAATCTCTCTGCTGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 544
DB 1042 -----CCGGATCCCGTAAACACACGCGCAAGGCCATTTGACTCAGGTTTGGT 995
QY 545 CGGTTCTCGGGGATGATCATGAGCTTGGGACCATATCAATATTGTTCAAGTTCAACA 604
DB 994 CAGCATGTCGCAATTTTCAATTAAGCGCATTTAGTTGCTTAAACACAGCCTCTTTGCCCT 935
QY 605 TGCCCTCAGACAGGAGCTTACCTGCTGCGGGAACCTCTGGAATTTCCATCGAGAT 664
DB 934 TAGCAAAAGTGGCTTGTGTTGGTATGATTTTGAAGTATCCGGAATTTCCCTTTAGT 875
QY 665 ATTTGCTCGTGGAGCGGCTCGCGCAAGTGGTGAGAAAGCAATGACGCCAAGACCAT 721
DB 874 AGCGATCTGATACAGGCTTCTGATAACGCTCGTATGCAATCAACCCGCGACCAT 818

; PRIOR APPLICATION NUMBER: US/60/235,863
 ; PRIOR FILING DATE: 2000-09-27
 ; NUMBER OF SEQ ID NOS: 2276
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1034
 ; LENGTH: 431
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-954-456-1034

Query Match 4.6%; Score 40; DB 10; Length 431;
 Best Local Similarity 53.1%; Pred. No. 0.0036;
 Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 211 CGCCCGCTGTCACAGCCCCAAGGTGGCGCTGGATCAACGTCGTCGCGCAGTTGTTGATG 270
 Db 2 CGCCCGCTGGTCAACCGCAGCGTCCCGAGTCGCTCTCGGGCGTGGTCTTGTATGAAG 61

Qy 271 ACCGATTGGCCATCAAACTGATTGATGGTTAGTTTTTCGCGGGTTTTTGAAATFCGGTG 330
 Db 62 TAGTGCCTGTCCTTGGCCCTCGATGCTGAAGTCAGGTTCTCCAGGTAGAAGCGCTTTGTTG 121

Qy 331 GCCTTCGCCCAAAATGTTGATGCCGCGCTCGTGGGAAATCT 370
 Db 122 AGCACGGCGCCACCTGTATGATCAGTCTCTCGTTGGCGATGT 161

RESULT 10
 US-09-738-626-1672/c
 ; Sequence 1672, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OKAZI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 1672
 ; LENGTH: 948
 ; TYPE: DNA
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1672

Query Match 4.4%; Score 38.2; DB 9; Length 948;
 Best Local Similarity 63.7%; Pred. No. 0.021;
 Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 634 GCGCGGAACCTCTGGAATTCATCGAGATATTTGTCGTGAGCAGCGCCTCGCGCAAGT 693
 Db 689 GCTCTGGAATCATGGGAAATTCGAGCGGTATTTAGCAGTCAGCACGCTTGCCCGCAGC 630

Qy 694 GGTGAGAAAGCAATGACGCCAAGACCAATTGT 724
 Db 629 GGAGCGCCAGCAAGAAATCCGACACCTTAGT 599

; ORGANISM: Corynebacterium glutamicum
US-09-738-626-439

Query Match 4.1%; Score 36; DB 9; Length 1068;
Best Local Similarity 53.6%; Pred. No. 0.12; Mismatches 0; Gaps 0;
Matches 75; Conservative 0; Indels 65; Indels 0; Gaps 0;

QY 146 CTGGCGGTTTCGGCGCAAGCTGATCTGGTTCGGTGGTTCGGCGCAGCAGCAT 205
DB 542 CCGCGCGCTGACCTGCTTCTTATCTACATCGCGCGCAGAGCTCCGCGAAGAAATTA 601
QY 206 TGTACGCGCGCTGTCAGCCCAAGTGTGGCGCTGGATCAAGTCTGTCGGCAGTTG 265
DB 602 TTTTCTGGCAGATGGGTTCCTCAACGGCTCGAATGGCGGCAGCTCAACGTGGTGTG 661
QY 266 TGATGACCGCATTTGGCCATC 285
DB 662 TGATCGTGAATCGGCTC 681

RESULT 14

US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 4.1%; Score 36; DB 9; Length 3309400;
Best Local Similarity 53.6%; Pred. No. 5.1;
Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 146 CTGGCGGTTTCGGCGCAAGCTGATCTGGTTCGGTGGTTCGGCGCAGCAGCAT 205
DB 414184 CCGCGCGCTGACCTGCTTCTTATCTACATCGCGCGCAGAGCTCCGCGAAGAAATTA 414243
QY 206 TGTACGCGCGCTGTCAGCCCAAGTGTGGCGCTGGATCAAGTCTGTCGGCAGTTG 265
DB 414244 TTTTCTGGCAGATGGTTCCTCAACGGCTCGAATGGCGGCAGCTCAACGTGGTGTG 414303
QY 266 TGATGACCGCATTTGGCCATC 285
DB 414304 TGATCGTGAATCGGCTC 414323

RESULT 15

US-09-815-242-7775/c
; Sequence 7775, Application US/09815242

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7775
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(996)
US-09-815-242-7775

Query Match 4.1%; Score 35.4; DB 10; Length 996;
Best Local Similarity 53.2%; Pred. No. 0.19;
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 470 TGGTCAAGTATCCGCGCGTACTCTCTTGTCTCGGCGAGCACCCTGCAAGCGCATCT 529
DB 847 TGGTCCCGGAATCGGATCAGGTCTATCGCCCTGCGCGAGCACCCTGCGCGAGT 788
QY 530 GCGCAAGTGACTGCCCGGCTTCTCGGCGATGTCTATTGAGCTTGGGACCATATCAATAT 589
DB 787 GCGAGGGTTTACGGCCCTTGGTGGCGCAGTTGCCCGACCTTGTCCACAGCTTCAGGT 728
QY 590 TGTTCAGTTCAACATGCCCT 610
DB 727 TCTTCGCGAAGTTCTCGCCCT 707

Search completed: April 27, 2003, 05:34:59
Job time : 3284.59 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:22:28 ; Search time 49.1862 Seconds
(without alignments)
5443.169 Million cell updates/sec

Title: US-09-105-117K-1_COPY_1421_2293

Perfect score: 873

Sequence: 1 gtcgataagcagcgggttg.....ctgtgagctctgacgcgtag 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*

2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*

3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*

4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*

5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71.6	8.2	5541	1	US-08-920-812-20
2	71.6	8.2	5541	1	US-08-920-827-20
3	71.6	8.2	5541	1	US-08-921-177-20
4	71.6	8.2	5541	1	US-08-362-577C-20
5	71.6	8.2	5541	2	US-08-920-828-20
6	57	6.5	15239	1	US-08-390-878-17
7	57	6.5	4403765	4	US-09-103-840A-2
8	57	6.5	4411529	4	US-09-103-840A-1
9	39.6	4.5	1935	2	US-08-492-027A-9
10	37.6	4.3	4403765	4	US-09-103-840A-2
11	37.6	4.3	4411529	4	US-09-103-840A-1
12	34.4	3.9	2048	3	US-08-776-246-1
13	34.4	3.9	2048	3	US-08-776-246-1
14	33.6	3.8	1681	4	US-09-434-288-7
15	33.6	3.8	8438	1	US-07-945-283-1
16	33.2	3.8	20235	3	US-07-642-734C-3
17	33.2	3.8	20235	3	US-08-439-009A-3
18	33	3.8	707	4	US-08-998-416-940
19	32.8	3.8	4508	5	PCT-US93-06251-34
20	32.6	3.7	421	3	US-08-782-480-7
21	32.6	3.7	421	4	US-08-954-211-7
22	32.6	3.7	421	4	US-09-005-167A-7
23	32.6	3.7	421	4	US-09-176-741B-7
24	32.4	3.7	1680	4	US-09-603-185-1
25	32.4	3.7	4689	4	US-09-105-537-34
26	32.4	3.7	36778	4	US-09-105-537-5
27	32.4	3.7	38506	3	US-09-320-878-19

28	32.2	3.7	1404	4	US-08-979-608A-10
29	32.2	3.7	2625	6	5457037-4
30	32.2	3.7	3336	6	5457037-1
31	32	3.7	1608	4	US-09-292-225-20
32	32	3.7	1608	4	US-09-292-225-22
33	32	3.7	1665	4	US-09-292-225-17
34	32	3.7	1665	4	US-09-292-225-19
35	32	3.7	1752	4	US-09-292-225-14
36	32	3.7	1752	4	US-09-292-225-16
37	32	3.7	1929	4	US-09-380-420C-1
38	31.6	3.6	1241	1	US-08-471-033-39
39	31.6	3.6	1241	1	US-08-471-033-42
40	31.6	3.6	1241	2	US-08-471-044-39
41	31.6	3.6	1241	2	US-08-471-044-42
42	31.6	3.6	1241	2	US-08-463-483A-39
43	31.6	3.6	1241	2	US-08-463-483A-42
44	31.6	3.6	1241	2	US-08-471-046A-39
45	31.6	3.6	1241	2	US-08-471-046A-42

ALIGNMENTS

RESULT 1
US-08-920-812-20
; Sequence 20, Application US/08920812
; Patent No. 5763188
; GENERAL INFORMATION:

APPLICANT: Ohno, Tsuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,812
FILING DATE: 29-AUG-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:

NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 5541 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: Clinical Isolate EC-625

US-08-920-812-20

QY 791 AATAGCTTGGCTGATGAATCAGAAAGCGGGCAGCCCTCTCCGCCATGAATCAGCCGCT 850
 Db 3253 TGTACGAAGGTTGATGAATTAACAGCGGAATTTTCCACTCGCGCAGCACTCAACCAATTT 3312
 QY 851 CGCTGTGAGCTCTGGACCGTA 872
 Db 3313 TTTGGCTCCGCTCTGGCGAGTA 3334

RESULT 3

US-08-921-177-20
 ; Sequence 20, Application US/08921177
 ; Patent No. 5798211
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/921,177
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,577
 ; FILING DATE: 27-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5541 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli
 ; STRAIN: Clinical Isolate EC-625
 ; US-08-921-177-20

Query Match 8.2%; Score 71.6; DB 1; Length 5541;
 Best Local Similarity 49.4%; Pred. No. 1.7e-11;
 Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;
 QY 203 CATTGTGACGCGCTGTGACCCCAAGGTGGCGCTGGATCAACGTCGTGGGCGAG 262
 Db 2687 CATTATCCATCTGACCTTTGTGCTCTCTGATTTGTGATGGGCTATCTTGGACCGACTGAC 2746
 QY 263 TTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGGTTAGTTTCGCGGTTTGG 322
 Db 2747 TCGTCATATCACTGATCTTATGCGGGCGGGTTCTCGCGCCGGTTATTAAACAGGTCAAT 2806
 QY 323 AATCGGTGGCCTTCGCCCCAAATGTTGATGCCGCGCTGTGGGAAATCTCATCGATCGCCT 382

Db 2807 TATCGGAAGACGCGCTGCCACAGATTTCAGCTCGCATCGCGGATATGCTGATCAATCTGCG 2866
 QY 383 CCAACTCGCGCTCAGAAAACCTCAAGTTTGTGAGTGAATCAAGGCTGTTGTCCAGCTGCT 442
 Db 2867 CCAGCTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2926
 QY 443 CAACTGACGAAGCACCACCAATCAATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
 Db 2927 CGCGCGC-TGGCACCACCAATCAATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2967
 QY 503 CGCGCAGCACCCATGCAAGCGCATCTGCGCAAGTGAATCTGCGCGCTTCTTGGGCGATGT 562
 Db 2968 CTTTCAGCAACCCAGCTTAAACGCCATTGCGCCATTGATTGTCACGCTGCTGTCGCAATTT 3027
 QY 563 CATTGAGCTTGGGACCATATCAATATTGTTCAAGTTTCAAGTCCCTCAGACAGGACT 622
 Db 3028 CATTCAATAAGTGTAGGCTGTTGAGGTTTGGCTTCCGTAAGCATTTTCGGCGCTCAGACCCAC 3087
 QY 623 TACCTTGGCTGGC-----GCGGGAACCTCTCTGGAATTCATCGAGATATTTCT 670
 Db 3088 GAATTTATCTCTTCAAGTCCGCTGCAATCTTGGGAATGCGTTGAGATATTTTC 3147
 QY 671 CGGTGAGCAGGCGCTTGGCAAGTGGTGAGAAAGCAATGACGCGCAAGACCATTTGTTGGCAG 730
 Db 3148 CGGTGAGCAATCCCTGAGCCAGAGGAGTAAGGCAATACAGCCCGCGGTTATTTGCA 3207
 QY 731 CTGACTGCAACAAAGTTCTCACCGTCTATCGCCGCTTCTCCACCCCAACGATTAAATGATGG 790
 Db 3208 GGGTATCCAGCA-----GCGCGCTTTTATCCACCCAGCGGTTTCAGTAAAT 3252
 QY 791 AATAGCTTGGCTGATGAATCAGAAAGCGGCGCAGCCCTCTCGCCATGAATCAGCGGCT 850
 Db 3253 TGTACGAAGTTGATGAATTAACAGCGGAATTTTCACTCGCGCAGCACTCAACCAATTT 3312
 QY 851 CGCTGTGAGCTCTCGACCGTA 872
 Db 3313 TTTGGCTCCGCTCTGGCGAGTA 3334

RESULT 4

US-08-362-577C-20
 ; Sequence 20, Application US/08362577C
 ; Patent No. 5807673
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/362,577C
 ; FILING DATE: 27-MAR-1995
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300

```

; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5541 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-625
; US-08-362-577C-20

Query Match      8.2%; Score 71.6; DB 1; Length 5541;
Best Local Similarity 49.4%; Pred. No. 1.7e-11;
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;

QY 203 CATTGTGACGGCCGCTGTCAGCCCAAGGTGTGGCGCTGGATCAACGTGCTGGCGAG 262
DB 2687 CATTATCCATCTGACGTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGACTGAC 2746
QY 263 TTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATGGTTAGTTTTCGCGGTTTGG 322
DB 2747 TCGTCATATCACTGATCTTATGCGGCGGGTTCTCGGCGCGGTTATTAAACAGGTCATT 2806
QY 323 AATCGGTGGCTTCGCGCAAAATGTTGATGCGCGCTGCTGGGAAATCTCATGATCGCCT 382
DB 2807 TATCGGAAGACGCGCTGCCACAGATTTCAGCTCGCCATCGCGCATATGCTGATCAATCTGCG 2866
QY 383 CCAACTCGGGCTCAGAAACTCCAAAGTTGTTGATGTAATCAAGCTGTTGTCAGCTGCT 442
DB 2867 CCAGCTCTCGGTGCTAAATGTCAGATTATTTCAGCGCTTCGACGTTCTCTCAAGTTGTC 2926
QY 443 CAACTGACGAAGCACCATAATCAATGCACTGCTCAGGTTATCCGCGCGGTACTCTCTTGTCT 502
DB 2927 CGCGCGGCTGGCACCAATCAATACCGAGCTCAGCGATCA-----T 2967
QY 503 CGCGCAGCACCATTGCAAGCGCCATCTGCGCAAGTGAAGTGCCTGCGGTTCTTGGCGGATGT 562
DB 2968 CTTTCAGCAACCAAGCTTAACGCCATTTGGGCCATTGTTGTCACGCTGCTGTGCCAATTT 3027
QY 563 CATTGAGCTTGGGACCAATATCAATATTTGATGTTGATGTTCAAGTCAATGCTGCGGCTCAGACGACT 622
DB 3028 CATTCAATAAGTGTAGGCTGTGAGGTTGCTTGGTAAGCAATTTTCGGGCTCAGACCAAC 3087
QY 623 TACCTGCTGCGC-----GCGGGAACCTCTGGAATTCATCGAGATATTGT 670
DB 3088 GAATTTATTCCTTTCAGATGCAATCGTGAATCTTGGGAAATGCGTTGAGATATTTC 3147
QY 671 CCGTGAGCAGCGCTTGGCAAGTGGTGAAGCAATGAGCGCAAGACCAATGTTGTTGGAG 730
DB 3148 CGGTGAGCAATCCCTGAGCAGAGAGTAAAGGCAATACAGCCCAAGCGGCTTATTGCA 3207
QY 731 CTGACTGCAACAGTTCTCAGGTCATCGCGGCTTCTTCCACCCAGCAATATGATGG 790
DB 3208 GGGTATCCAGCA-----GCGCGCTTTTATTCACCCAGGCTTCAGTAAAT 3252
QY 791 AATAGCTTGGCTGATGAATCAGAGCGGCGAGCCCTCTCTCGCCATGAATCAGCGCGCT 850
DB 3253 TGTACGAGGTTGATGAATTAACGCGGAATTTCCACTCGCGGAGCAATCAACCATTT 3312
QY 851 CCGCTGTGAGCTTGGACCGTA 872
DB 3313 TTTGCTCGGCTGTGGCGAGTA 3334

RESULT 5
US-08-920-828-20
; Sequence 20, Application US/08920828
; Patent No. 5853998
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya

```

```

; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,828
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5541 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-625
; US-08-920-828-20

```

```

Query Match      8.2%; Score 71.6; DB 2; Length 5541;
Best Local Similarity 49.4%; Pred. No. 1.7e-11;
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;

QY 203 CATTGTGACGGCCGCTGTCAGCCCAAGGTGTGGCGCTGGATCAACGTGCTGGCGAG 262
DB 2687 CATTATCCATCTGACGTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGACTGAC 2746
QY 263 TTGTGATGACCGCATTCGCGCATCAAACTGATGTTGATGGTTAGTTTTCGCGGTTTGG 322
DB 2747 TCGTCATATCACTGATCTTATGCGGCGGGTTCTCGGCGCGGTTATTAAACAGGTCATT 2806
QY 323 AATCGGTGGCTTCGCGCAAAATGTTGATGCGCGCTGCTGGGAAATCTCATGATCGCCT 382
DB 2807 TATCGGAAGACGCGCTGCCACAGATTTCAGCTCGCCATCGCGCATATGCTGATCAATCTGCG 2866
QY 383 CCAACTCGGGCTCAGAAACTCCAAAGTTGTTGATGTAATCAAGCTGTTGTCAGCTGCT 442
DB 2867 CCAGCTCTCGGTGCTAAATGTCAGATTATTTCAGCGCTTCGACGTTCTCTCAAGTTGTC 2926
QY 443 CAACTGACGAAGCACCATAATCAATGCACTGCTCAGGTTATCCGCGCGGTACTCTCTTGTCT 502
DB 2927 CGCGCGGCTGGCACCAATCAATACCGAGCTCAGCGATCA-----T 2967
QY 503 CGCGCAGCACCATTGCAAGCGCCATCTGCGCAAGTGAAGTGCCTGCGGTTCTTGGCGGATGT 562
DB 2968 CTTTCAGCAACCAAGCTTAACGCCATTTGGGCCATTGTTGTCACGCTGCTGTGCCAATTT 3027
QY 563 CATTGAGCTTGGGACCAATATCAATATTTGATGTTGATGTTCAAGTCAATGCTGCGGCTCAGACGACT 622

```

Db 3028 CATTAATAGTGTAGGCTGTTGAGGTTGCTTTCGGTAAGCAATTTTCGGCGTCAGACCC 3087
 QY 623 TACCTTGGCTGGC-----GGGGAACCTCTGAATTCATCAGATATTTGT 670
 Db 3088 GAATCTTAATTCCTTTCACGATGATCGTGAATCTTTCGGGAATGCGGTGAGATATTTTC 3147
 QY 671 CCGTAGCAGCGCTCGCGAAGTGGTGAAGAAAGCAATGAGCCCAAGACCACTGTGTCAG 730
 Db 3148 CGGTACGAACTCTGAGCCAGAGGAGTAAGGCATACAGCCCAAGCGCGTATTTTGCA 3207
 QY 731 CTGACTGCAACAAGTTCTCACCGTCAATCGCCCGGTTCCCTCAACCCAAAGTAATGATGG 790
 Db 3208 GGGTATCCAGCA-----GGCCGCTTTTATCCACCCAGCGGTTTCAGTAAAT 3252
 QY 791 AATAGCTTGCTGATGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAACTCAGCCGCT 850
 Db 3253 TGACGAAGTTGATGAATTAACAGCGGAATTTTCCACTCGCGCAGCAACTCAACCAATTT 3312
 QY 851 CCCTGTGAGCTCTGGACCGTA 872
 Db 3313 TTTGCGTCCGCTCTGGCGATA 3334

RESULT 6

US-08-390-878-17/c
 ; Sequence 17, Application US/08390878
 ; Patent No. 5700683

GENERAL INFORMATION:

; APPLICANT: Stover, Charles K.
 ; APPLICANT: Mahairas, Gregory G.
 ; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
 ; NUMBER OF SEQUENCES: 18
 ; CORRESPONDENCE ADDRESSES:

; ADDRESSEE: Townsend and Townsend Kourile and Crew
 ; STREET: One Market Plaza, Steuart Street Tower, 20th
 ; FLOOR:
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94105

COMPUTER READABLE FORM:

; MEDIUM TYPE: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; FILING DATE: 17-FEB-1995
 ; APPLICATION NUMBER: US/08/390,878
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

; NAME: Hunter, Tom
 ; REGISTRATION NUMBER: 38,498
 ; REFERENCE/DOCKET NUMBER: 15371A-17
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415/543/9600
 ; TELEFAX: 415/543/5043

INFORMATION FOR SEQ ID NO: 17:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 15239 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-390-878-17

Query Match 6.5%; Score 57; DB 1; Length 15239;
 Best Local Similarity 53.1%; Pred. No. 6.4e-07;
 Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCCATGTTGATGGAATCGTCTGACCTGGTTGAACCCGAATGCGTATTTGGACGCGT 85
 Db 4792 AGGTCTGCTGACCTGTGCGGCATTCACGTTCTCAACCCACACAGCTCTACCTCGACACCG 4733

QY 86 TTGTGTTTATCGGCGCGTTCGGCGCAATACGGCGACACCGGAGTGGATTTTCGCCG 145
 Db 4732 TGTGTTGCTAGCGCGCTGGCCAAAGACACAGCA-----CCAGCGCTGGCTGTTTCGGCC 4676
 QY 146 CTGGCGCGTTTCGGCAAGCCTGATCTGTTCCCGTGGTGGGTTTCGGCGCAGCAGCAT 205
 Db 4675 TCGGCGCGGTACAGCAGTGGGTATGTTCCGCCACCCCTCGGGTTCGGAGCCGCGGT 4616
 QY 206 TGTACGCCCGCTGTCCAGCCCCAAGGTGTGGCGTGGATCAACGTCGTCGTCGAGATTG 265
 Db 4615 TCGCGCGGCTGTTCACCAACCCCGGCTCGTGAGAAATCCTCGACGCGCTGATCGCGGTCA 4556
 QY 266 TGATGACCCCATTTGGCCATCAAACTGATTTCA 298
 Db 4555 TGATGTTGCGTGGGAATCTCGCTGACCGTGA 4523

RESULT 7

US-09-103-840A-2
 ; Sequence 2, Application US/09103840A
 ; Patent No. 6294328

GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; FILE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 2

TYPE: DNA

LENGTH: 4403765

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 6.5%; Score 57; DB 4; Length 4403765;

Best Local Similarity 53.1%; Pred. No. 9.1e-06;

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCCATGTTGATGGAATCGTCTGACCTGGTTGAACCCGAATGCGTATTTGGACGCGT 85

Db 2227633 AGTCTGCTGACCTGTGCGGCATTCAGTTCTCAACCCACACGTTACTCTGACACCG 2227692

QY 86 TTGTGTTTATCGGCGCGTTCGGCGCAATACGGCGACACCGGAGTGGATTTTCGCCG 145

Db 2227693 TCGTGTGCTAGCGCGCTGGCCAAAGACACAGCA-----CCAGCGCTGGCTGTTTCGGCC 2227749

QY 146 CTGGCGCGTTTCGGCAAGCCTGATCTGTTCCCGTGGTGGGTTTCGGCGCAGCAGCAT 205

Db 2227750 TCGGCGCGGTCAAGCCAGTGGGTATGTTTCGGCACCCCTCGGGTTCGGAGCCGCGGT 2227809

QY 206 TGTACGCCCGCTGTCCAGCCCCAAGGTGTGGCGTGGATCAACGTCGTCGTCGAGATTG 265

Db 2227810 TCGCGCGGCTGTTCACCAACCCCGGCTCGTGAGAAATCCTCGACGCGCTGATCGCGGTCA 2227869

QY 266 TGATGACCCCATTTGGCCATCAAACTGATTTCA 298

Db 2227870 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 2227902

RESULT 8

US-09-103-840A-1
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328

```

; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
; US-09-103-840A-1

Query Match 6.5%; Score 57; DB 4; Length 4411529;
Best Local Similarity 53.1%; Pred. No. 9.1e-06;
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCGATGTTGATGGCAATCTGCTGACCTGGTTGAACCGAATGCGTATTGTGACGCGT 85
DB 2230334 AGGTCTCTGGTGACCTGTGCGGCAATTCAGTTCTCAACCCACACGCTACTCTGACACCG 2230393

QY 86 TTGTGTTTATCGGCGGCGGTTCGGCGGCAATACGGCGACACCGGCGGTGGATTTTCGCGC 145
DB 2230394 TCGTGTGTCTAGCGCGCGCTGGCCACGAGCACACGGA---CCAGCGCTGGCTGTTTCGCC 2230450

QY 146 CTGCGCGGCTTCGCGGCAAGCCCTGATCTGTGTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT 205
DB 2230451 TCGCGCGGCTCACAGCCAGTGCGGTATGTTTCGCCACCTTCGGGTTTCGAGCGCGCGGT 2230510

QY 206 TGTCAAGCCCGCTGTTCAGACCCCAAGGTGTGGCGCTGGATCAACGTGCTGTGGCAGTTG 265
DB 2230511 TGGCGGGGCTGTTTCAACCAACCCCGGCTGTGGAGAAATCTTCGACGGCCTGATCGCGTCA 2230570

QY 266 TGATGACCGCAATGGCCATCAAACTGATGTTGA 298
DB 2230571 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 2230603

```

RESULT 9
US-08-492-027A-9/c
; Sequence 9, Application US/08492027A
; Patent No. 5912333
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shoichi
; APPLICANT: Burnell, James N
; TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch and Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/492.027A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr, Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0760-206

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1935 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1638
; US-08-492-027A-9

Query Match 4.5%; Score 39.6; DB 2; Length 1935;
Best Local Similarity 49.1%; Pred.No. 0.038;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 222 CAGCCCCAAGGTGTGGCGCTGGATCAACGTGCTGCGCAGTTGTGATGACCGCATTTGGC 281
Db 1476 CGGCACCGAGGCGTGCTTTCTTTACCTTCATCTTGGCAATGAAGCCGATCTTTAACCCA 1417
Qy 282 CATCAAACTGATGTGATGGTTAGTTTTCGGGGTTTTTGGAAATCGGTGGCCTTCGGCCCA 341
Db 1416 GTCCCTCAGCAAGTGAAGGTGTAGGCTGCGCCGTTTCCTTGGAGGGAGAGAGCGGCCCT 1357
Qy 342 AATGTTGATCGCGCGTCGTGGGAAATCTCATCGATCGCTCCAACTCGCGGTGAGAAA 401
Db 1356 GATGCCACCGCAGCATGTATGCGCAATGACACAGAGACCTCCACCTTGAGGGCGCACAC 1297
Qy 402 CTCCAAGTTGTTGAGTGAATCAAGGCTGTTGTCC 435
Db 1296 AGCTACTCGATGCGGACCGCATGCCGGTGTAC 1263

RESULT 10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

```

	Query Match	4.3%;	Score 37.6;	DB 4;	Length 4403765;
	Best Local Similarity	47.5%;	Pred. No. 5.4;		
	Matches 112;	Conservative 0;	Mismatches 124;	Indels 0;	Gaps 0;
Qy	25	AAGCCCATGTTTATGGCAATCGTCTGACCTGGTTGAACCCGAATGCGTATTTGGACGCG	84		
Db	720443	AAGGACCTGCGCCTTGGCCGTCTGAGGGGGCGGCGGAGCTGCCGGGCCCCCGCTCGGT	720384		
Qy	85	TTTGTGTTTATCGCGGGCGTTCGGCGCGCAATACGGCGGACACCGGACGGTGTGATTTTCGCC	144		
Db	720383	GGCTGCTGTTTCACTTGTCAAAGGCGCGGGAGCGAGCAATGTTTCTGGGTTCAACGACACACAC	720324		

QY 145 GCTGGCGGTTTCGGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGGCGCAGCA 204
DB 720323 CGCTCGACGATCGAGGACCTGCTGGCGGGATTCGCTGGCGGTTTCTTCGCGCGGG 720264
QY 205 TTGTACGCCCGCTGTCAGCCCAAGTGTGGCGCTGATCAAGCTGCTGGC 260
DB 720263 GAGATCGGCCCGGTCGGCGGCCACACGCTGTGACCGGTTTACCGGCTGATGGC 720208

RESULT 11

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 4.3%; Score 37.6; DB 4; Length 4411529;
Best Local Similarity 47.5%; Pred. No. 5.4; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 124;
QY 25 AAGCCCATGTTGATGGCAATGCTGCTGACCTGGTTGAACCCGAATGCGTATTGGACGG 84
DB 719010 AAGGACCTGGCGCTGGCTGAGCGGGCGCGGAGTCCCGGCGCCCGCTCGT 718951
QY 85 TTGTGTATTATCGCGCGGCTGGCGCGGATACGCGGACACCGGATTTTCGCC 144
DB 718950 GGGCTGCTGTTTCACTGCAACGGCGGCGGACGAAATGTTCCGGGTCAACGACCA 718891
QY 145 GCTGGCGGTTTCGGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGGCGCAGCA 204
DB 718890 CGCTCGACGATCGAGGACCTGCTGGCGGGATTCGCTGGCGGTTTTCGCGCGGG 718831
QY 205 TTGTACGCCCGCTGTCAGCCCAAGTGTGGCGCTGGATCAACGCTGCTGGC 260
DB 718830 GAGATCGGCCCGGTCGGCGGCCACACGCTGTGACCGGTTTACCGGCTGATGGC 718775

RESULT 12

US-08-776-246-1/c
; Sequence 1, Application US/08776246
; Patent No. 6004550
; GENERAL INFORMATION:
; APPLICANT: Springer, Caroline J
; APPLICANT: Marais, Richard
; TITLE OF INVENTION: Intracellular expression of carboxypeptidase
; TITLE OF INVENTION: G2 in enzyme produg therapy
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderyhe
; STREET: 1100 No. 6004550th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,246
FILING DATE: 30-JAN-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/01783
FILING DATE: 27-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9415167.7
FILING DATE: 27-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Arthur R. Crawford
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 620-19
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2048 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-776-246-1

Query Match 3.9%; Score 34.4; DB 3; Length 2048;
Best Local Similarity 48.0%; Pred. No. 1.4;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 461 TCAATGCACTGCTCAGGCTATCGCGCCCTACTCTCTTGTCTGGCGAGCACCACCA 520
DB 1950 TCGATGATCGCTGCGCGCTCGGCTCGGTGAGGAGGCGCGGATCCGGCCCT 1891
QY 521 GCGCCATTCGCCAAGTGAATGCTCCCGCGTTCTTGGCGGATGTCATTGAGCTTGGCGACCA 580
DB 1890 GCGCCAGCGCTCGAGCCACTGCTTCTTCTGCTGGCGTTGCGGTAGCGCATGAGGATGG 1831
QY 581 TATCATATTTGTCAGTTCAACATGCCCTCAGACAGGAGACTTACCCTGGCTGGCGCGGG 640
DB 1830 CGTTGACGCGGCGAGTTGGTTCAGCTGATGGCGGTGCTGTGCGCGCTGCTGGCGCGCGCA 1771
QY 641 AACCTCTCTGGAATTCATCGAGAT 664
DB 1770 TCTCTTCCAGCACCAGCGCAGGT 1747

RESULT 13

US-08-776-251-1/c
; Sequence 1, Application US/08776251
; Patent No. 6025340
; GENERAL INFORMATION:
; APPLICANT: Springer, Caroline J
; APPLICANT: Marais, Richard
; TITLE OF INVENTION: Surface expression of enzyme in gene directed produg therapy
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderyhe
; STREET: 1100 No. 6025340th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,251
FILING DATE: 31-JAN-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/01782
FILING DATE: 27-JUL-1995

PRIOR APPLICATION DATA: GB 9415167.7
FILING DATE: 27-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Arthur R. Crawford
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 620-20
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2048 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-776-251-1

Query Match 3.9%; Score 34.4; DB 3; Length 2048;
Best Local Similarity 48.0%; Pred. No. 1.4;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 461 TCAATGCACTGGTCACGGTATCCGCGCTACTCTCTCTGTCGCGCAGCACCCTGCA 520
DB 1950 TCGATGATCGCTCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 1891
QY 521 CGCGCATCTGGCAAGTACTCCCGGCTTCTCTGGCGATGTCTATGAGCTTGGGACCA 580
DB 1890 CGCGCAGCGCTCGAGCACTGCTTCTCTGCTCGCGGTTGCCGTAGCGATGAGGATGG 1831
QY 581 TATCAATATTGTTTACCTGTTCAACATGCCCTCAGACAGGACTTACCTGCTGGCGGGG 640
DB 1830 CGTTGACGGGCGATGTTGTTACGCTGATGGCGGTGCTGTCGCCCTCGCGCGCGCA 1771
QY 641 AACCTCTGGAATTCATCGAGAT 664
DB 1770 TCTCTCCAGCACCAGCGCGAGGT 1747

RESULT 14
US-09-434-288-7
Sequence 7, Application US/09434288
Patent No. 6303767
GENERAL INFORMATION:
APPLICANT: Betlach C., Melanie
APPLICANT: McDaniel, Robert
TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
TITLE OF INVENTION: CONSTRUCTS THEREFOR
FILE REFERENCE: 30062-20030.00
CURRENT APPLICATION NUMBER: US/09/434,288
CURRENT FILING DATE: 1999-11-05
PRIOR APPLICATION NUMBER: 60/107,093
PRIOR FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 7
LENGTH: 1681
TYPE: DNA
ORGANISM: Streptomyces narbonensis
US-09-434-288-7

Query Match 3.8%; Score 33.6; DB 4; Length 1681;
Best Local Similarity 49.4%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 87; Conservative 0; Mismatches 89;

QY 115 TACGGGACACCGACCGTGGATTTTCGCCGCTGGCGGTTTCGGGCAAGCTGATCTGG 174
DB 1493 TACTGCCGTCGGCGGGTTCCTCCAGCAGCGGGGAGTTTCGACGCGGCTTCTTCGGG 1552
QY 175 TTCCGCTGGTGGTTCGGGCGGACGAGTATTCACGCCGCTGTTCAGCCCAAGTG 234
DB 1553 ATCTGCGCGCGAGGCGCTCGCATGACCCGCGAGCAGCGGCTGTCCCTCACACCGCG 1612
QY 235 TGGCGCTGGATCAACGTCGTGTGCGAGTTGTGATGACCGGATTTGGCCATCAACT 290

DB 1613 TGGGAGGCGATCGACGCGGGCATCGACCGGCGCTGAAGGCGAGCGGCT 1668

RESULT 15
US-07-945-283-1
Sequence 1, Application US/07945283
Patent No. 5352596
GENERAL INFORMATION:
APPLICANT: Cheung, Andrew K.
APPLICANT: Wesley, Ronald D.
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants
TITLE OF INVENTION: Involving The BP0 and LLT Genes
NUMBER OF SEQUENCES: 7.
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis P. Ribando
STREET: 1815 No. 5352596th University Street
CITY: Peoria
STATE: IL
COUNTRY: USA
ZIP: 61604
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/945,283
FILING DATE: 19920911
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Ribando, Curtis P.
REGISTRATION NUMBER: 27976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 309-685-4011 ext. 513
TELEFAX: 309-685-4128
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8438 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudorabies virus
FEATURE:
NAME/KEY: CDS
LOCATION: 622..6495
FEATURE:
NAME/KEY: variation
LOCATION: replace(1099, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1267, "t")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1381, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1566, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(7010, "g")
US-07-945-283-1

Query Match 3.8%; Score 33.6; DB 1; Length 8438;
Best Local Similarity 51.3%; Pred. No. 4.7; Indels 0; Gaps 0;
Matches 78; Conservative 0; Mismatches 74;

QY 337 GCCCAATGTTGATCGCGGCTCGTGGGAAATCTCATCGATCGCTCCCAACTCGGGTCA 396
DB 6434 GCCCGACCGGGGACCGCGGCGCGGCGGCGGCTCGTCTCTCTCTCTCTCTCTCTCG 6493

Qy 397 GAAACTCCAAAGTTGTTGAGTGAATCAAGGCTGTTGTCCAGCTGCTCAACTGACGAAGCA 456
Db 6494 TCTAGCACACGATCTCGCCCGAGCCCGCGGGGGTCCGCTGCTGCTGGGCGGAGGA 6553
Qy 457 CCAATCAATGCACTGGTCAAGGTATCCGCGCC 488
Db 6554 GGACGGGGCGGCTCGTGGCTCCGCGCGCGGC 6585

Search completed: April 27, 2003, 07:20:35
Job time : 12814.2 secs

Result No.	Query No.	Score	Query			DB	ID	Description
			Match	Length	%			
1	350	29.4	211	1	Q0EC5A	hypothetical 23k p		
2	347	29.1	211	2	A0807A	probable membrane		
3	343	28.8	211	2	B91103	hypothetical prote		
4	340	28.5	211	2	B85948	hypothetical prote		
5	335	28.1	205	2	A0G012	probable LysE type		
6	332	27.9	200	2	D83100	probable transport		
7	326.5	27.4	202	2	C97472	probable membrane		
8	326.5	27.4	202	2	A02690	LysE family transp		
9	319	26.8	201	2	C07044	hypothetical prote		
10	318	26.7	199	2	H07566	hypothetical prote		
11	302.5	25.4	204	2	A03411	transporter, LysE		
12	245.5	20.6	211	2	B82318	LysE/YggA family p		
13	238	20.0	211	2	S57940	YggA protein homol		
14	176.5	14.8	220	2	D98938	conserved hypoteth		
15	173.5	14.6	210	2	F84609	conserved hypoteth		
16	170.5	14.3	210	2	B71906	hypothetical prote		
17	164	13.8	205	2	B99854	conserved hypoteth		
18	154.5	13.0	205	2	C97219	uncharacterized co		
19	153.5	12.9	213	2	E97789	hypothetical prote		
20	151.5	12.7	208	1	B69086	conserved hypoteth		
21	148	12.4	200	2	G83703	hypothetical prote		
22	136.5	11.5	213	2	F83444	hypothetical prote		
23	132	11.1	210	2	E87252	efflux protein, Ly		
24	130	10.9	210	2	D84016	hypothetical prote		
25	120.5	10.1	208	2	G87305	efflux protein, Ly		
26	119	10.0	620	2	F82449	potassium uptake p		
27	117.5	9.9	197	2	B83280	hypothetical prote		
28	115.5	9.7	216	2	AH3203	RhtB family transp		
29	114.5	9.6	208	2	G84086	dihydrodipicolinat		

```

RESULT 2
AH0874
Probable membrane transport protein STY3222 [imported] - Salmonella enterica subsp. ente
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0874
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AH0874
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <PAR>
A:Crossa-references: GB:AL513382; PIDN:CAD02896.1; PID:g16504149; GSPDB:GN00176
C:Genetics:
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 29.1%; Score 347; DB 2; Length 211;
Best Local Similarity 33.6%; Pred. No. 3e-24;
Matches 79; Conservative 42; Mismatches 80; Indels 34; Gaps 4;

QY 3 IMEITFGLLGASLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTLGVDL 62
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MISYFGVALGAAMILPLGPQNAFVNMQIRQYHLMIALCALSLVLSAGIFGSGA 60
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 63 LSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAAPQIIEETPTVDDTPLGSSAV 122
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 LMQSPWLLALVTWGGVAFLLWYGFAGKATAMSSNLEL----- 98

QY 123 ATDTRNRVRVEVSDKQVWVKPMLMAIVLTWLPNAYLDAFVIGVGQAQYG-DTGRWI 181
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 -----ASAEVLKQGRW-KIATMLAVTWNPHVYLDTFVVLGSLGQLAMEPKRW- 147

QY 182 FAAGAPASLIWFLVGFGAALSRPLSSPKVWRWVNVVAVMTALAIAKL 236
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 148 FALGTISASFLWFLGLAALAAWLAAPRLRTAKSQRIINLVGVVWMLIAFQLAREG 202
: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
B91103
hypothetical protein ECs3794 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: B91103
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <HAY>
A:Crossa-references: GB:BA000007; PIDN:BAE37217.1; PID:g13363266; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 28.8%; Score 343; DB 2; Length 211;
Best Local Similarity 35.2%; Pred. No. 6.9e-24;
Matches 80; Conservative 37; Mismatches 76; Indels 34; Gaps 4;

QY 7 FITGLLGASLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTLGVDL 66
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 YFOGLALGAAMILPLGPQNAFVNMQIRQYHLMIALCALSLVLSAGIFGSGALLMQ 64
: : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 67 APIVLDMRWGGIAYLLWFAVMAAKDAMTKVKAAPQIIEETPTVDDTPLGSSAVATDT 126
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 SPWLLALVTWGGVAFLLWYGFAGKATAMSSNLEL----- 98

QY 127 RNRVRVEVSDKQVWVKPMLMAIVLTWLPNAYLDAFVIGVGQAQYG-DTGRWIFAAG 185
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 -----ASAEVLKQGRW-KIATMLAVTWNPHVYLDTFVVLGSLGQLDVEPKRW-FALG 151

QY 186 AFAASLIWFLVGFGAALSRPLSSPKVWRWVNVVAVMTALAIAKL 232
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 TISASFLWFLGLAALAAWLAAPRLRTAKSQRIINLVGVVWMLIAFQL 198

RESULT 4
E85948
hypothetical protein yggA [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85948
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-211 <STO>
A:Crossa-references: GB:AE005174; NID:g12517455; PIDN:AAG58049.1; GSPDB:GN00145; UMGP:242
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: yggA
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 28.5%; Score 340; DB 2; Length 211;
Best Local Similarity 34.5%; Pred. No. 1.3e-23;
Matches 78; Conservative 37; Mismatches 79; Indels 33; Gaps 2;

QY 7 FITGLLGASLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTLGVDL 66
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 YFOGLALGAAMILPLGPQNAFVNMQIRQYHLMIALCALSLVLSAGIFGSGALLMQ 64
: : : : : : : : : : : : : : : : : : : : : : : : : :

QY 67 APIVLDMRWGGIAYLLWFAVMAAKDAMTKVKAAPQIIEETPTVDDTPLGSSAVATDT 126
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 SPWLLALVTWGGVAFLLWYGFAGKATAMSSNLEL----- 98

QY 127 RNRVRVEVSDKQVWVKPMLMAIVLTWLPNAYLDAFVIGVGQAQYDGTGRWIFAAGA 186
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 99 -----ASAEVLKQGRW-KIATMLAVTWNPHVYLDTFVVLGSLGQLDVXPKRWFAAGT 152

QY 187 FAASLIWFLVGFGAALSRPLSSPKVWRWVNVVAVMTALAIAKL 232
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 ISASFLWFLGLAALAAWLAAPRLRTAKSQRIINLVGVVWMLIAFQL 198

RESULT 5
AG0112
probable lyse type translocator YPO0918 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AG0112
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <KUR>
A:Crossa-references: GB:AL590842; PIDN:CAC89762.1; PID:g15978989; GSPDB:GN00175
C:Genetics:

```

A:Gene: YPO0918
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 28.1%; Score 335; DB 2; Length 205;
Best Local Similarity 33.1%; Pred. No. 3.6e-23;
Matches 80; Conservative 44; Mismatches 70; Indels 48; Gaps 6;

QY 3 IMEIFTGLLGASLLSSTGPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVDL 62
DB 1 MLAVYLHGFTLSAAILPLGPQNVFVWQGIKQHLMSALCALSDIILICAGIFGSSA 60

QY 63 LSNAAPVLDIMRWGGIAYLLWP---AVMAAKDAMTKVAPQIIEETEPTVDDTPLGG 119
DB 61 LLSRSPLLLALVTWGGVAFPLMWYWGALMAA-----WRG 94

QY 120 SAVATD---TRNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQ-Y 174
DB 95 DGVASATSVTQGRWILLVT-----LAVTLNPNHVIDTTFVVLGSLGQLL 141

QY 175 GDTGRWIFPAAGAPAAALIFPLVFGGAALSRPLSSPKVWRWVNVVAVMTALAIKML 234
DB 142 PDIRPW-FALGAVTASIVWFFALALLAALSPWLNRPVQRIINLFPVGMGFIAFQLAR 200

QY 235 MG 236
DB 201 QG 202

RESULT 6
DB3100
Probable transporter PA4365 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83100
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83100
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-200 <STO>
A:Cross-references: GB:AE004852; GB:AE004091; NID:g9950587; PIDN:AAG07753.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4365
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 27.9%; Score 332; DB 2; Length 200;
Best Local Similarity 33.5%; Pred. No. 6.5e-23;
Matches 77; Conservative 42; Mismatches 77; Indels 34; Gaps 4;

QY 7 FITGLLGASLLSSTGPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLLSNA 66
DB 5 YLNGILVAAGLIAIGAQAQNAFVLAQSLRREHLSVAALCVFCDAVLVSLGVFLAKLILE 64

QY 67 APVLDIMRWGGIAYLLWPFAVMAAKDAMTKVAPQIIEETEPTVDDTPLGGSAVAT 126
DB 65 NPTLIAIARWGGIAFLTWGLKALLRALR-----PDAL---GNAETGP 105

QY 127 RNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQYGDTRWIFPAAG 186
DB 106 RSRKAV-----LLAALAVTLNPNHVIDTTFVVLGSLGQAQAPG--AYALGA 150

QY 187 FAASLIWFLVFGGAALSRPLSSPKVWRWVNVVAVMTALAIKMLMG 236
DB 151 ASASLWFFPALGAALWAPLWLPATWELLDLIMVAAMMLGMAQLIFRG 200

RESULT 7
C97472

probable membrane transport protein. (AL357613) [imported] - Agrobacterium tumefaciens (C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: C97472
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: C97472
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK66732.1; PID:gl5155924; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C 1690
A:Map position: circular chromosome
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 27.4%; Score 326.5; DB 2; Length 202;
Best Local Similarity 31.6%; Pred. No. 2.1e-22;
Matches 73; Conservative 42; Mismatches 83; Indels 33; Gaps 4;

QY 4 MEIFTGLLGASLLSSTGPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLL 63
DB 3 IQIIFTGLTMGLSLIVAIGAQAQNAFVLAQSLRSHVFAVCATCAISDALLIMVGVFGFQRI 62

QY 64 SNAAPVLDIMRWGGIAYLLWPFAVMAAKDAMTKVAPQIIEETEPTVDDTPLGSAVA 123
DB 63 SAIMPALDPIIMRYAGAAFLIMYGAKSLYSALRSS-EVLSVAERREAS----- 108

QY 124 TDTNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQYGDTRWIFPA 183
DB 109 -----LW-QTALAICLALTPLNPHVIDTTFVVLGTLTISTQPPGFEK-TFA 149

QY 184 AGAFAASLIWFLVFGGAALSRPLSSPKVWRWVNVVAVMTALAIKML 234
DB 150 AGAATGSLFFPFSGLYGARWLRIPEKESAWRLISGLVIAITWATAFKLVM 200

RESULT 8
AG2690
Lyse family transporter lyseE [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AG2690
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2690
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41941.1; PID:gl7739308; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: lyseE
A:Map position: circular chromosome
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 27.4%; Score 326.5; DB 2; Length 202;
Best Local Similarity 31.6%; Pred. No. 2.1e-22;
Matches 73; Conservative 42; Mismatches 83; Indels 33; Gaps 4;

QY 4 MEIFTGLLGASLLSSTGPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLL 63
DB 3 IQIIFTGLTMGLSLIVAIGAQAQNAFVLAQSLRSHVFAVCATCAISDALLIMVGVFGFQRI 62

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:20:37 ; Search time 11 Seconds
(without alignments)
889.856 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVMEIFITGLLGALLLS.....INWVAVVMTALAIAKIMLMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	98.9	233	1 LYSE CORGL	P94633 corynebacte
2	350	29.4	211	1 YGGA ECOLI	P11667 escherichia
3	319	26.8	201	1 Y488 MYCTU	Q11154 mycobacteri
4	318	26.7	199	1 YJ86 MYCTU	Q10871 mycobacteri
5	310	26.0	206	1 YGGA AERSA	P70775 aeromonas s
6	254	21.3	225	1 YGGA AERHY	P52047 aeromonas h
7	105	8.8	206	1 RHTC ECOLI	P27846 escherichia
8	101	8.5	212	1 YEAS ECOLI	P76249 escherichia
9	93.5	7.9	452	1 PTCC ECOLI	P17334 escherichia
10	92.5	7.8	465	1 Y093 RHIME	O87394 rhizobium m
11	92	7.7	223	1 YAHN ECOLI	P75693 escherichia
12	91	7.6	206	1 RHTC SALTY	Q91877 salmonella
13	89	7.5	944	1 CHS2 NEUCR	P30589 neurospora
14	88.5	7.4	3137	1 CA36 CHICK	P15989 gallus gall
15	88	7.4	206	1 RHTC SALTI	Q823b3 salmonella
16	87.5	7.3	443	1 DCUA HELPY	O25425 helicobacte
17	87	7.3	295	1 YBIF SALTY	P74867 salmonella
18	86.5	7.3	195	1 YBIF ECOLI	P38101 escherichia
19	86	7.2	295	1 YBIF ECOLI	P36545 escherichia
20	86	7.2	351	1 MRAY METTH	O26830 methanococc
21	85	7.1	302	1 Y091 METJA	Q57556 methanococc
22	84.5	7.1	443	1 DCUA HELPU	Q921c0 helicobacte
23	84	7.1	482	1 YF1G BACSU	P54723 bacillus su
24	84	7.1	685	1 PHUB SALTY	O87656 salmonella
25	83	7.0	216	1 YBFT PSEAR	P38102 pseudomonas
26	82.5	6.9	3491	1 ERYI SACER	Q03131 saccharopol
27	81.5	6.8	253	1 CY5Z SALTI	Q824w3 salmonella
28	81.5	6.8	742	1 DHET ACEAC	P18278 acetobacter
29	81	6.8	622	1 KUP ECOLI	P30016 escherichia
30	81	6.8	676	1 CCWF RHIME	P45404 rhizobium m
31	80.5	6.8	253	1 CY5Z ECOLI	P12610 escherichia
32	80.5	6.8	253	1 CY5Z SALTY	P12673 salmonella
33	80	6.7	206	1 RHTB ECOLI	P27847 escherichia

RESULT 1

ID	LYSE CORGL	STANDARD;	PRT;	233 AA.
AC	P94633;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Lysine exporter protein.			
GN	LYSE OR CGL1262.			
OS	Corynebacterium glutamicum (Brevibacterium flavum)			
OC	Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;			
OC	Actinomycetales; Corynebacterineae; Corynebacteriaceae;			
OC	Corynebacterium.			
OX	NCBI_TaxID=1718;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND FUNCTION.			
RC	STRAIN=RI27;			
RX	MEDLINE=97126810; PubMed=89711704;			
RA	Vrljic M.M., Sahm H., Eggeling L.;			
RT	"A new type of transporter with a new type of cellular function: L-			
RT	lysine export from Corynebacterium glutamicum.";			
RL	Mol. Microbiol. 22:815-826(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;			
RA	Nakagawa S.;			
RT	"Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";			
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: INVOLVED IN THE EFFLUX OF EXCESS OF L-LYSINE. THIS IS			
CC	NECESSARY TO CONTROL THE INTRACELLULAR L-LYSINE LEVEL.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.			
CC	-!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; X96471; CAA65324.2; .			
DR	EMBL; AF005277; BAB98655.1; ALT_INIT.			
DR	InterPro; IPR001123; LyseE.			
DR	InterPro; IPR004777; Lyse_exporter.			
DR	Pfam; PF01810; LyseE; 1.			
DR	TIGRFAMs; TIGR00948; 2a75; 1.			
DR	Transport; Transmembrane; Inner membrane.			
KW	TRANSMEM 3			
FT	TRANSMEM 35			
FT	TRANSMEM 55			
FT	TRANSMEM 66			
FT	TRANSMEM 86			
FT	TRANSMEM 144			
FT	TRANSMEM 164			
FT	TRANSMEM 177			
FT	TRANSMEM 197			
FT	TRANSMEM 213			
FT	TRANSMEM 233			
FT	TRANSMEM 253			
FT	TRANSMEM 273			
FT	TRANSMEM 293			
FT	TRANSMEM 313			
FT	TRANSMEM 333			
FT	TRANSMEM 353			
FT	TRANSMEM 373			
FT	TRANSMEM 393			
FT	TRANSMEM 413			
FT	TRANSMEM 433			
FT	TRANSMEM 453			
FT	TRANSMEM 473			
FT	TRANSMEM 493			
FT	TRANSMEM 513			
FT	TRANSMEM 533			
FT	TRANSMEM 553			
FT	TRANSMEM 573			
FT	TRANSMEM 593			
FT	TRANSMEM 613			
FT	TRANSMEM 633			
FT	TRANSMEM 653			
FT	TRANSMEM 673			
FT	TRANSMEM 693			
FT	TRANSMEM 713			
FT	TRANSMEM 733			
FT	TRANSMEM 753			
FT	TRANSMEM 773			
FT	TRANSMEM 793			
FT	TRANSMEM 813			
FT	TRANSMEM 833			
FT	TRANSMEM 853			
FT	TRANSMEM 873			
FT	TRANSMEM 893			
FT	TRANSMEM 913			
FT	TRANSMEM 933			
FT	TRANSMEM 953			
FT	TRANSMEM 973			
FT	TRANSMEM 993			
FT	TRANSMEM 1013			
FT	TRANSMEM 1033			
FT	TRANSMEM 1053			
FT	TRANSMEM 1073			
FT	TRANSMEM 1093			
FT	TRANSMEM 1113			
FT	TRANSMEM 1133			
FT	TRANSMEM 1153			
FT	TRANSMEM 1173			
FT	TRANSMEM 1193			
FT	TRANSMEM 1213			
FT	TRANSMEM 1233			
FT	TRANSMEM 1253			
FT	TRANSMEM 1273			
FT	TRANSMEM 1293			
FT	TRANSMEM 1313			
FT	TRANSMEM 1333			
FT	TRANSMEM 1353			
FT	TRANSMEM 1373			
FT	TRANSMEM 1393			
FT	TRANSMEM 1413			
FT	TRANSMEM 1433			
FT	TRANSMEM 1453			
FT	TRANSMEM 1473			
FT	TRANSMEM 1493			
FT	TRANSMEM 1513			
FT	TRANSMEM 1533			
FT	TRANSMEM 1553			
FT	TRANSMEM 1573			
FT	TRANSMEM 1593			
FT	TRANSMEM 1613			
FT	TRANSMEM 1633			
FT	TRANSMEM 1653			
FT	TRANSMEM 1673			
FT	TRANSMEM 1693			
FT	TRANSMEM 1713			
FT	TRANSMEM 1733			
FT	TRANSMEM 1753			
FT	TRANSMEM 1773			
FT	TRANSMEM 1793			
FT	TRANSMEM 1813			
FT	TRANSMEM 1833			
FT	TRANSMEM 1853			
FT	TRANSMEM 1873			
FT	TRANSMEM 1893			
FT	TRANSMEM 1913			
FT	TRANSMEM 1933			
FT	TRANSMEM 1953			
FT	TRANSMEM 1973			
FT	TRANSMEM 1993			
FT	TRANSMEM 2013			
FT	TRANSMEM 2033			
FT	TRANSMEM 2053			
FT	TRANSMEM 2073			
FT	TRANSMEM 2093			
FT	TRANSMEM 2113			
FT	TRANSMEM 2133			
FT	TRANSMEM 2153			
FT	TRANSMEM 2173			
FT	TRANSMEM 2193			
FT	TRANSMEM 2213			
FT	TRANSMEM 2233			
FT	TRANSMEM 2253			
FT	TRANSMEM 2273			
FT	TRANSMEM 2293			
FT	TRANSMEM 2313			
FT	TRANSMEM 2333			
FT	TRANSMEM 2353			
FT	TRANSMEM 2373			
FT	TRANSMEM 2393			
FT	TRANSMEM 2413			
FT	TRANSMEM 2433			
FT	TRANSMEM 2453			
FT	TRANSMEM 2473			
FT	TRANSMEM 2493			
FT	TRANSMEM 2513			
FT	TRANSMEM 2533			
FT	TRANSMEM 2553			
FT	TRANSMEM 2573			
FT	TRANSMEM 2593			
FT	TRANSMEM 2613			
FT	TRANSMEM 2633			
FT	TRANSMEM 2653			
FT	TRANSMEM 2673			
FT	TRANSMEM 2693			
FT	TRANSMEM 2713			
FT	TRANSMEM 2733			
FT	TRANSMEM 2753			
FT	TRANSMEM 2773			
FT	TRANSMEM 2793			
FT	TRANSMEM 2813			
FT	TRANSMEM 2833			
FT	TRANSMEM 2853			
FT	TRANSMEM 2873			
FT	TRANSMEM 2893			
FT	TRANSMEM 2913			
FT	TRANSMEM 2933			
FT	TRANSMEM 2953			
FT	TRANSMEM 2973			
FT	TRANSMEM 2993			
FT	TRANSMEM 3013			
FT	TRANSMEM 3033			
FT	TRANSMEM 3053			
FT	TRANSMEM 3073			
FT	TRANSMEM 3093			
FT	TRANSMEM 3113			
FT	TRANSMEM 3133			
FT	TRANSMEM 3153			
FT	TRANSMEM 3173			
FT	TRANSMEM 3193			
FT	TRANSMEM 3213			
FT	TRANSMEM 3233			
FT	TRANSMEM 3253			
FT	TRANSMEM 3273			
FT	TRANSMEM 3293			
FT	TRANSMEM 3313			
FT	TRANSMEM 3333			
FT	TRANSMEM 3353			
FT	TRANSMEM 3373			
FT	TRANSMEM 3393			
FT	TRANSMEM 3413			
FT	TRANSMEM 3433			
FT	TRANSMEM 3453			
FT	TRANSMEM 3473			
FT	TRANSMEM 3493			
FT	TRANSMEM 3513			
FT	TRANSMEM 3533			
FT	TRANSMEM 3553			
FT	TRANSMEM 3573			
FT	TRANSMEM 3593			
FT	TRANSMEM 3613			
FT	TRANSMEM 3633			
FT	TRANSMEM 3653			
FT	TRANSMEM 3673			
FT	TRANSMEM 3693			
FT	TRANSMEM 3713			
FT	TRANSMEM 3733			
FT	TRANSMEM 3753			
FT	TRANSMEM 3773			
FT	TRANSMEM 3793			
FT	TRANSMEM 3813			
FT	TRANSMEM 3833			
FT	TRANSMEM 3853			
FT	TRANSMEM 3873			
FT	TRANSMEM 3893			
FT	TRANSMEM 3913			
FT	TRANSMEM 3933			
FT	TRANSMEM 3953			
FT	TRANSMEM 3973			
FT	TRANSMEM 3993			
FT	TRANSMEM 4013			
FT	TRANSMEM 4033			

Query Match 98.9%; Score 1178; DB 1; Length 233;
 Best Local Similarity 100.0%; Pred. No. 3.6e-94; Indels 0; Gaps 0;
 Matches 233; Conservative 0; Mismatches 0;

QY 4 MEIFITGLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 63
 DB 1 MEIFITGLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 60

QY 64 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKVPAQIIEETPTVPDDTPLGSSAVA 123
 DB 61 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKVPAQIIEETPTVPDDTPLGSSAVA 120

QY 124 TDTNRNRVSVSDVKQVWVKPMLMAVLTLNPNAYLDAFVIGVGGAQYG-DTGRWIFAAG 185
 DB 121 TDTNRNRVSVSDVKQVWVKPMLMAVLTLNPNAYLDAFVIGVGGAQYG-DTGRWIFA 180

QY 184 AGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKLMG 236
 DB 181 AGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKLMG 233

RESULT 2
 YGGA_ECOLI
 ID YGGA_ECOLI STANDARD; PRT; 211 AA.
 AC P11667;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ygga.
 GN YGGA OR B2923.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=9278503;
 RA Blatner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glaser J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 RN [2]
 RP SEQUENCE OF 15-211 FROM N.A.
 RC STRAIN=K12 / CS520;
 RX MEDLINE=89313302; PubMed=2546007;
 RA Alefounder P.R., Baldwin S.A., Perham S.A., Short N.J.;
 RT "Identification, molecular cloning and sequence analysis of a gene
 RT cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-
 RT phosphoglycerate kinase and a putative second glyceraldehyde 3-
 RT phosphate dehydrogenase of Escherichia coli.";
 RL Mol. Microbiol. 3:723-732 (1989).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U28377; AAA69090.1;
 EMBL; A5000375; AAC75960.1;
 EMBL; X14436; CA32607.1;
 PIR; S04736; Q8ECSA.
 BioGene; EGI1159; YGGA.
 InterPro; IPR001123; Lyse.
 InterPro; IPR004777; Lys_exporter.
 Pfam; PF01810; Lyse; 1.

DR TIGRFAMs: TIGR00948; 2a75; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 37 57 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 SQ SEQUENCE 211 AA; 23175 MW; 2DAFE27B6A9B822 CRC64;

Query Match 29.4%; Score 350; DB 1; Length 211;
 Best Local Similarity 35.7%; Pred. No. 3.6e-23;
 Matches 81; Conservative 37; Mismatches 75; Indels 34; Gaps 4;

QY 7 FITGLLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 66
 DB 5 YFQGLGAAMTILPLGPQNAFVMAKQIRQYHIMIALCAISDLVLCAGIFGSSALLMQ 64

QY 67 APVLDIMRWGGIAYLLWFAVMAAKDAMTKVPAQIIEETPTVPDDTPLGSSAVATDT 126
 DB 65 SPWLLALVTWGVAFLLWYFGAFKPTAMSSNIEL----- 98

QY 127 RNRVRVSVSDVKQVWVKPMLMAVLTLNPNAYLDAFVIGVGGAQYG-DTGRWIFAAG 185
 DB 99 -----ASAEVWKQGRW-KIATMLAVTLNPNAYLDAFVIGVGGAQYG-DTGRWIFAAG 185

QY 186 AFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKLMG 232
 DB 152 TISAGFLWFFGLAALWAPLRITAKAQRILNVVGVGMVFIALQL 198

RESULT 3
 Y488_MYCTU
 ID Y488_MYCTU STANDARD; PRT; 201 AA.
 AC Q11154;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV0488.
 GN RV0488 OR MT0507 OR MTCY2069.14.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RN Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; Z77162; CAB00949.1; --
 EMBL; AE006952; AAK44730.1; --
 TIGR; MT0507; --
 TubercuList; RV0488; --
 InterPro; IPR001123; LysE.
 InterPro; IPR004777; Lys_exporter.
 Pfam; PF01810; LysE; 1.
 TIGRPFAMs; TIGR00948; 2a75; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 25 45 POTENTIAL.
 FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 104 124 POTENTIAL.
 FT TRANSMEM 133 153 POTENTIAL.
 FT TRANSMEM 169 189 POTENTIAL.
 SQ SEQUENCE 201 AA; 20951 MW; E198975DF088E5E4 CRC64;

Query Match 26.8%; Score 319; DB 1; Length 201;

Best Local Similarity 32.4%; Pred. No. 1.6e-20;
 Matches 71; Conservative 50; Mismatches 64; Indels 34; Gaps 4;

QY 16 SLLLSIGPQNVLVKQIKREGIALLVCLISDVFLFIAGTIGVLDLSNAAPVLVDIMR 75
 DB 3 TLKVALGFPQNAFLVRQGIIRREYVLIVVLCGADGALIAAGVGFGFAALIHAFNMTLVAR 62
 QY 76 WGGIAYLLWFVAVMAAKDAMTNKVEAPQIIEETPTVDDTPLGGSAVATDRNRVREVS 135
 DB 63 FGGAFLIGYALLAARNW-----RPSGLVPSES--GPAAL----- 96
 QY 136 VDKQVWVKPMLMAVLVTWLNPNAYLDAFVPIGVGQAGYDGTGRWTFPAAGAFASLIWFP 195
 DB 97 -----IGVVQMLVTFPLNPHVYLDVTVLIGALANESDL-RWFFGAGAAASVVWFA 148
 QY 196 LVFGAAALSRPLSSPKVWVWVWVAVMTALAIAKML 234
 DB 149 VLGFSGRLQFPFATPAWRILDAALVATMIGVAVVVLV 187

RESULT 4

YJ86 MYCTU STANDARD; PRT; 199 AA.
 ID YJ86 MYCTU STANDARD; PRT; 199 AA.
 AC Q10871;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV1986
 GN RV1986 OR MT2040 OR MTCY39.33C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544 (1998).

RN

RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -|- SIMILARITY: BELONGS TO THE LYS_E/YGGA FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

EMBL; Z74025; CA989398.1; --
 EMBL; AE007056; AAK46315.1; --
 TIGR; MT2040; --
 TubercuList; RV1986; --
 InterPro; IPR001123; LysE.
 InterPro; IPR004777; Lys_exporter.
 Pfam; PF01810; LysE; 1.
 TIGRPFAMs; TIGR00948; 2a75; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 6 26 POTENTIAL.
 FT TRANSMEM 42 62 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 116 136 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 178 198 POTENTIAL.
 SQ SEQUENCE 199 AA; 20775 MW; 6F330132DDCFD0FF CRC64;

Query Match

Best Local Similarity 26.7%; Score 318; DB 1; Length 199;
 Matches 70; Conservative 44; Mismatches 77; Indels 34; Gaps 3;

QY 8 ITGLLIGASLLSIGPQNVLVKQIKREGIALLVCLISDVFLFIAGTIGVLDLSNA 67
 DB 6 VVGFACFTLIAIGAQAFLVRQGIQREHVPVVALCTVSDIVLIAGIAGFGALIGAH 65
 QY 68 PIVLDIMRWGGIAYLLWFVAVMAAKDAMTNKVEAPQIIEETPTVDDTPLGGSAVATDTR 127
 DB 66 PRALNVVKGGAFLIGYGLLAARRARFVALIP-----SGATPVR 106
 QY 128 NRVREVSVDKORVWVKPMLMAVLVTWLNPNAYLDAFVPIGVGQAGYDGTGRWTFPAAGAF 187
 DB 107 -----LAEVLVTCAAFPLNPHVYLDVTVLIGALANESDQ-RWLFGLGAV 151
 QY 188 AASLIWFLVFGGAALSRPLSSPKVWVWVWVAVMTALAIAKML 232
 DB 152 TASAVWFATLGFAGRLGLFTNPGSWRLDGLIAVMVVALGISL 196

RESULT 5

YGGA AERSA
 ID YGGA AERSA STANDARD; PRT; 206 AA.
 AC P70775;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein YGGA.
 GN YGGA.
 OS Aeromonas salmonicida.
 OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
 OC Aeromonas.
 OX NCBI_TaxID=645;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NCIMB 1102;
RX  MEDLINE=97431471; PubMed=9286976;
RA  Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RA  Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.;
RT  "Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
RT  identification of the LuxRI homologs AhyRI and AsaRI and their
RT  cognate N-acylhomoserine lactone signal molecules.";
RL  J. Bacteriol. 179:5271-5281(1997).
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC  -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U6570019.1; ALT_INIT.
DR  InterPro; IPR001123; Lyse.
DR  InterPro; IPR004777; Lys_exporter.
DR  Pfam; PF01810; Lyse; 1.
DR  TIGRFAMs; TIGR00948; 2a75; 1.
KW  Hypothetical protein; Transmembrane.
FT  TRANSMEM 1 21 POTENTIAL.
FT  TRANSMEM 37 57 POTENTIAL.
FT  TRANSMEM 65 85 POTENTIAL.
FT  TRANSMEM 116 136 POTENTIAL.
FT  TRANSMEM 148 168 POTENTIAL.
FT  TRANSMEM 185 205 POTENTIAL.
SQ  SEQUENCE 206 AA; 21505 MW; DIC2C492CDA0179A CRC64;

Query Match      26.0%; Score 310; DB 1; Length 206;
Best Local Similarity 31.9%; Pred. No. 9.5e-20;
Matches 72; Conservative 43; Mismatches 77; Indels 34; Gaps 3;

QY  8 ITGLLGASLLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGVDLLSNAA 67
DB  6 LQGFITGLAMIIPIGAQNAPVLSRGIRHNHLLATLCCCLDLVLIGVFGGANLAAAS 65
QY  68 PIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETEPTVPDDTPLGGSVAATDTR 127
DB  66 PIGLALLTWGGVFLCWFGRGIRSLRSWQGO-----GAALADSPR 104
QY  128 NRVVRVSVDRQVWVKPML-MAIVLTWLNPNAYLDFAVFVGGAQYGDGTGRWIFAAGA 186
DB  105 -----LMGVKSVLAWTLGVTLNPNHYLDLTLMLLGSFGSQFAEPLRPAAGA 152
QY  187 FAASLWFLPVLGFGAALSRPLSSPKVWRWVNVAVVMTALAIKL 232
DB  153 MLASLVNWFYSLAFGAALSPWLARGVQQAIDTIVGLIMLGLALQL 198

RESULT 6
YGGA_AERHY
ID  YGGA_AERHY STANDARD; PRT; 225 AA.
AC  P52047;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Hypothetical 24.5 kDa protein in ahyr-cdnp intergenic region.
OS  Aeromonas hydrophila.
OC  Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC  Aeromonas.
OX  NCBI_TaxID=644;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=AI;
RX  MEDLINE=97431471; PubMed=9286976;
RA  Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,

```

```

RA  Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.;
RT  "Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
RT  identification of the LuxRI homologs AhyRI and AsaRI and their
RT  cognate N-acylhomoserine lactone signal molecules.";
RL  J. Bacteriol. 179:5271-5281(1997).
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC  -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; X89469; CAA61655.1; ALT_INIT.
DR  InterPro; IPR001123; Lyse.
DR  InterPro; IPR004777; Lys_exporter.
DR  Pfam; PF01810; Lyse; 1.
DR  TIGRFAMs; TIGR00948; 2a75; 1.
KW  Hypothetical protein; Transmembrane.
FT  TRANSMEM 1 21 POTENTIAL.
FT  TRANSMEM 37 57 POTENTIAL.
FT  TRANSMEM 65 85 POTENTIAL.
FT  TRANSMEM 116 136 POTENTIAL.
FT  TRANSMEM 150 170 POTENTIAL.
SQ  SEQUENCE 225 AA; 24482 MW; 172DB104473B0809 CRC64;

Query Match      21.3%; Score 254; DB 1; Length 225;
Best Local Similarity 26.9%; Pred. No. 6.6e-15;
Matches 61; Conservative 38; Mismatches 74; Indels 54; Gaps 4;

QY  8 ITGLLGASLLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGVDLLSNAA 67
DB  6 LQGFITGLAMIIPIGAQNAPVLSRGIRHNHLLATLCCCLDLVLIGVFGGANLAAAS 65
QY  68 PIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETEPTVPDDTPLGGSVA 123
DB  66 PIGLALLTWGGVFLCWFGRGIRSLRSWQGOAKLADSPQ-----MGVRSV- 111
QY  124 TDTNRVRVSVDRQVWVKPMLMAIVLTWLNPNAYLDFAVFVGGAQYGDGTGRWIFA 183
DB  112 -----LAWTLGVTLNPNHYLDLTLMLLGSFGSQFAEPLRPA 149
QY  184 AGAFAASLWFLPVLGFGAALSRPLSSPK-----VWRW 216
DB  150 AVAMLASLVNWFYSLAFGAALSPWLARGVQYKSLILLILLVSPCWGW 196

RESULT 7
RHTC_ECOLI
ID  RHTC_ECOLI STANDARD; PRT; 206 AA.
AC  P27846;
DT  01-AUG-1992 (Rel. 23, Created)
DT  15-JUL-1998 (Rel. 36, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Threonine efflux protein.
GN  RHTC OR B3823 OR Z5344 OR ECS4753.
OS  Escherichia coli, and
OS  Escherichia coli O157:H7.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562, 83334;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RX  MEDLINE=92358234; PubMed=1379743;
RA  Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT  "Analysis of the Escherichia coli genome: DNA sequence of the region
RT  from 84.5 to 86.5 minutes.";
RL  Science 257:771-778(1992).

```

DR	Pfam; PF01810; LysE; 1.
DR	TIGRFAMS; TIGR00949; 2A76; 1.
KW	Transport; Transmembrane; Complete proteome.
FT	TRANSMEM 1 21 POTENTIAL.
FT	TRANSMEM 44 64 POTENTIAL.
FT	TRANSMEM 67 87 POTENTIAL.
FT	TRANSMEM 150 173 POTENTIAL.
SQ	SEQUENCE 206 AA; 22474 MW; F64017878CC6D50D CRC64;

Query Match		8.8%; Score 105; DB 1; Length 206;
Best Local Similarity		20.4%; Pred. No. 0.037;
Matches 51; Conservative 36; Mismatches 99; Indels 64; Gaps		

Qy	3 IMEIFITGLLGSALLSIGQPONVLVIKQIKR---EGLIAYLLVFAVMAAKDMTNK---VEAPOIIBETEP	159
Dd	1 MLMLFLTAMVHVIALMSPGPDFFFVSQTAVSRKEAMVGULGTICGVVMWAGIA-LLG 59	59
Qy	60 VDILLSNAAPIVLDIMRW-----GGIAYLLWFPAVMAAKDMTNK---VEAPOIIBETEP	109
Dd	60 LHL-----IIERKAWLHFLTMVGGLYLCNMGYQLRGALKKEAVSAPAQV-----	106
Qy	110 TVPDDTPLGGSAAVATDTRNRRVRVEVSDQRVWVKPMALVLTWLNPNAYLDAFVPFG	169
Dd	107 -----ELAKGRS-----FLKLTLNLANPK-----AIIFYGS	134
Qy	170 -----VGAOYGDTGRIFAFAGAFAASLIWFPLVGFCAALSRPLSPKVKWINVVAVV	224
Dd	135 VFSLFVGDNGVTGARWGIFALIIVEITLWFTTVVASLFALPOMERGQRLAKWIDGFAGAL	194
Qy	225 MTALAIAKML 234	
Dd	195 FAGFGIHLLI 204	

RESULT 8	
ID	YEAS_ECOLI STANDARD; PRT; 212 AA.
AC	P76249; O07971; O07969;
DT	15-JUL-1998 (Rel. 36, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Hypothetical protein yeas.
GN	YEAS OR B1798.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_taxid=562;
[1]	SEQUENCE FROM N.A.
RN	STRAIN=K12 / MG1655;
RC	MEDLINE=97426617; PubMed=9278503;
RX	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley J.M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.:
RT	"The complete genome sequence of Escherichia coli K-12.";
RL	Science 277:1453-1474 (1997).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12;
RX	MEDLINE=97251358; PubMed=9097040;
RA	Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA	Kasai H., Kimura S., Kikagawa M., Kikagawa M., Makino K., Mikami T.,
RA	Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA	Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA	Yamamoto Y., Horuchi T.:
RT	"A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT	corresponding to the 40.1-50.0 min region on the linkage map.";
RL	RNA Res. 3:379-392(1996).
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC	-1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC	CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AB000274; AAC74868.1; --
 CC EMBL; D90823; BAA15593.1; --
 CC EMBL; D90824; BAA15602.1; --
 CC EcoGene; EGI3505; yeas.
 CC InterPro; IPR001123; 1; LysE.
 CC Pfam; PF01810; LysE; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 CC FT TRANSMEM 12 32 POTENTIAL.
 CC FT TRANSMEM 49 69 POTENTIAL.
 CC FT TRANSMEM 71 91 POTENTIAL.
 CC FT TRANSMEM 122 142 POTENTIAL.
 CC FT TRANSMEM 153 173 POTENTIAL.
 CC FT TRANSMEM 188 208 POTENTIAL.
 CC SQ SEQUENCE 212 AA; 23200 MW; 342E0DF348C9AD9A CRC64;
 Query Match 8.5%; Score 101; DB 1; Length 212;
 Best Local Similarity 22.3%; Pred. No. 0.083;
 Matches 52; Conservative 28; Mismatches 69; Indels 84; Gaps 9;
 QY 12 LIGASLLST-GPQNVILVKGKIK---REGLIAVLVLCLSDVFLFTAGTGLVGLDLSNAA 67
 Db 14 LVGAIFVLVPGNTLFLVKNSSGMKGGLAACGVFGDAVLMFLAWA-GVATLIKTT 72
 QY 68 PIVLDIMRWGGIAYLLW-----FAVMAAKDAMTNKVEAPQIIEETPTVPDPTLGGSA 121
 Db 73 PILFNIVRYLGAFLYLVGSKLVATLKGKNS-----EAKSDEPOYGAI 116
 QY 122 VADTNRNRVVEVDKQVRVWVPMALVILTWLNPNA---YLDFAVFVGCGAQQVGDGTG 178
 Db 117 -----FKRALISLUTNPKAILFYVSFFQFIDVNAHTGIS 152
 QY 179 RWIFAA-----CAFAA-----SLIWPLVGVGA 201
 Db 153 FFILATLELVSPCYLSFLISGAFVQYIRTKKLAKVGNLIGLMFVGFAA 205

RESULT 9

ID PTCC ECOLI STANDARD; PRT; 452 AA.
 AC P17334; P76212; P77333; P76907;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE PTS system, cellobiose-specific IIC component (BIIC-Cel) (Cellobiose-
 DE permease IIC component) (Phosphotransferase enzyme II, C component).
 GN CELB OR B1737.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90185127; PubMed=2179047;
 RA Parker L.L., Hall B.G.;
 RT "Characterization and nucleotide sequence of the cryptic cel operon
 RT of Escherichia coli K12.";
 RL Genetics 124:455-471 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 28.0-40.1 min region on the linkage map.";
 RL DNA Res. 3:363-377 (1996).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=91227627; PubMed=2092358;
 RA Reizer J., Reizer A., Saier M.H. Jr.;
 RT "The cellobiose permease of Escherichia coli consists of three
 RT proteins and is homologous to the lactose permease of Staphylococcus
 RT aureus.";
 RL Res. Microbiol. 141:1061-1067 (1990).
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
 CC -TRANSPORT SYSTEM. THE IIC DOMAINS CONTAIN THE SUGAR BINDING SITE
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
 CC THE SUGAR.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
 CC -!- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X52890; CAA37070.1; --
 CC EMBL; M64438; AAA23547.1; --
 CC EMBL; AB000268; AAC74807.1; --
 CC EMBL; D90816; BAA15518.1; --
 CC EMBL; D90817; BAA15526.1; --
 CC EMBL; D90818; BAA15532.1; --
 CC PIR; S10871; S10871.
 CC EcoGene; EGI0141; celB.
 CC InterPro; IPR004796; Cello_pts_IIC.
 CC InterPro; IPR003352; PTS_EIIC.
 CC InterPro; IPR004501; Pts.
 CC Pfam; PF02378; PTS_EIIC_1.
 CC TIGRfams; TIGR00359; Cello_pts_IIC; 1.
 CC TIGRfams; TIGR00410; pts; 1.
 CC KW Phosphotransferase system; Sugar transport; Transmembrane;
 KW Inner membrane; Complete proteome.
 FT TRANSMEM 35 55 POTENTIAL.
 FT TRANSMEM 86 106 POTENTIAL.
 FT TRANSMEM 114 134 POTENTIAL.
 FT TRANSMEM 147 167 POTENTIAL.
 FT TRANSMEM 187 207 POTENTIAL.
 FT TRANSMEM 229 249 POTENTIAL.
 FT TRANSMEM 300 320 POTENTIAL.
 FT TRANSMEM 344 364 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 FT CONFLICT 190 205 LIPGPFIILVMGIIAW -> FNSRLYISFRDGYCL (IN
 REF. 1).
 FT CONFLICT 233 235 VGM -> WL (IN REF. 1).
 FT CONFLICT 240 265 FVPLWFSGHGAALTALDNGMTP -> LSTALVLRHA
 ACADRTGQRHYDA (IN REF. 1).

[illegible]

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; AF233324; RAP33433.1; -;
DR	ENBL; AS008884; AAL22803.1; -;
DR	StyGene; SG????; rhtc.
DR	InterPro; IPR004778; Homoser_Thr_eff.
DR	InterPro; IPR001123; Lyse.
DR	Pfam; PF01810; Lyse; 1.
DR	TIGRFAMs; TIGR00949; 2A76; 1.
KW	Transport; Transmembrane; Complete proteome.
FT	TRANSMEM 1 21 POTENTIAL.
FT	TRANSMEM 44 64 POTENTIAL.
FT	TRANSMEM 67 87 POTENTIAL.
FT	TRANSMEM 150 173 POTENTIAL.
SQ	SEQUENCE 206 AA; 22498 MW; 080477853FC2733F CRC64;

```

Query Match          7.6%; Score 91; DB 1; Length 206;
Best Local Similarity 20.3%; Pred. No. 0.58;
Matches 50; Conservative 34; Mismatches 106; Indels 56; Gaps 8;

QY      3  IMEIFITGLLGLASLLSIGPONVLVIKQIGIKREGLIAVLLVCLISDFVFIAGT--LGV 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1  MMLPFTVAMVHIVALMSPGDFFFVSQTSVSRKSEAMMVGILGITCGVMVWAGVALLGL 60

QY      61  DLLSNRAPIVLDIRMGGIIAYLWPAWMAAKDAMTWK---VEAPQIIETEPTVPDDTPL 117
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61  HLIERKAWLHTTIMVGGGYLCWVGVMRGALKKQDAAASGPHI----- 106

QY      118  GGSAAVATDTRNRVRVESVDKQRVWVKPMLMAIVLTWLNPNAYLDAPVFTGG-----VGA 172
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      107  ---ELAQSGRS-----FLKGLLTNLNSPK---AIIYFGSVFSLFVGD 142

QY      173  QYGDTRGIWFAAGAF-----ASLIWPLVGFGAALSRPLSSPKWVRWVNVVAVVMTAL 228
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      143  NVGAARW-----GIFALITLETIAFTVVASLFAIPKMRGYYQRLAKWIDGPAGALFAGF 198

QY      229  AIKML 234
      : : :
Db      199  GIHLII 204

```

```

Db      199 GIBLII 204
|:::
RESULT 13
CHS2_NEUCR
ID CHS2_NEUCR STANDARD; PRT; 944 AA.
AC P30589;
DT DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chitin synthase 2 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl transferase 2).
DE CHS-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95039879; PubMed=7952169;
RA Din A.B., Yarden O.;
RT "The Neurospora crassa chs-2 gene encodes a non-essential chitin synthase.";
RL Microbiology 140:2189-2197(1994).
RN [2]
RP SEQUENCE OF 250-438 FROM N.A.
RX MEDLINE=92115692; PubMed=1731323;
RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
PA Robbins P.W.
```



```

FT DOMAIN 1239 1410 WFA 7.
FT DOMAIN 1441 1621 WFA 8.
FT DOMAIN 1641 1814 WFA 9.
FT DOMAIN 1840 2029 WFA 10.
FT DOMAIN 2407 2587 WFA 11.
FT DOMAIN 2625 2821 WFA 12.
FT DOMAIN 2945 3043 FIBRONECTIN TYPE-III.
FT DOMAIN 3068 3137 BPTI/KUNITZ INHIBITOR.
FT SITE 2166 2172 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2254 2259 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2308 2309 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2045 2047 CELL ATTACHMENT SITE.
FT SITE 2153 2155 CELL ATTACHMENT SITE.
FT SITE 2159 2161 CELL ATTACHMENT SITE.
FT ACT_SITE 3082 3083 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 3072 3122 BY SIMILARITY.
FT DISULFID 3081 3105 BY SIMILARITY.
FT DISULFID 3097 3118 BY SIMILARITY.
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2084 2084 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2436 2436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2563 2563 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2581 2581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2683 2683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2867 2867 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3137 AA; 339595 MW; ECB428578B536357 CRC64;

Query Match 7.4%; Score 88.5; DB 1; Length 3137;
Best Local Similarity 25.4%; Pred. No. 18;
Matches 46; Conservative 22; Mismatches 74; Indels 39; Gaps 6;

QY 14 GASLLLSIGPQNVLVKQIK-----REGLIAVLAVCLISDVFLFIAGTIGVLL----- 63
Db 337 GGSRIEAVPQILVLSGGSSDIREGLAVKQ-----ASIFSISGVNADSAELQOIA 392

QY 64 --SNAAPITVLDIMRWGGIAYLLWFAVMAAKDAMTKNVEAPOIIEETEPTVPDTPPL---G 118
Db 393 TDGSFATLDIRNLALRELLLNIVGVQORLI-LIEAPTIVTEVIVNKDILVFLDG 451

QY 119 GSAVATDTRNRVRVEVSVDKORVWVKPMLMAIVLTWLNPNAYLDAFVFIGGQYGDGTG 178
Db 452 STALGTGPFNSIRDFAKIVQRLVGVGPDLIQAV-----AQYADTV 492

QY 179 R 179
Db 493 R 493

RESULT 15
RHTC_SALTI
ID_RHTC_SALTI STANDARD; PRT; 206 AA.
AC Q8Z3B3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Threonine efflux protein.
GN RHTC OR STY3600;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jørgensen K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;

```

```

RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627278; CAD07933.1; -.
DR InterPro; IPR004778; Homoser_Thr_eff.
DR Pfam; PF01810; LyseE; 1.
DR TIGRfams; TIGR00349; 2A76; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22480 MW; E9DFCDF4B42D46A0 CRC64;

Query Match 7.4%; Score 88; DB 1; Length 206;
Best Local Similarity 19.9%; Pred. No. 1.1;
Matches 49; Conservative 35; Mismatches 106; Indels 56; Gaps 8;

QY 3 IMEIPITGILLGASLLSIGPQNVLVKQIKREGLIAVLAVCLISDVFLFIAGT---LCV 60
Db 1 MLMLFTVAMVHVALMSPGDPFFVSQTAVSRSRKEAMMGVLGITCGVMWVAGVALLGL 60

QY 61 DLLSNAAPITVLDIMRWGGIAYLLWFAVMAAKDAMTKN---VEAPOIIEETEPTVPDTPPL 117
Db 61 HLIEKMAWHLTIIVGGGLYLCWGYQMLRGALKQDAAASPHI----- 106

QY 118 GGSVATDTRNRVRVEVSVDKORVWVKPMLMAIVLTWLNPNAYLDAFVFIGG---VGA 172
Db 107 ---ELAQSGRS-----FLKGLLTNLSNPK---AIIYFGSVSLFVGD 142

QY 173 QYGDTCRWIFPAAGAYF-----ASLIWPLVFGGAALSRPLSPKVRWVWVAVVMTAL 228
Db 143 NVGAAARW---GIFALITLETIANFTVVASLFALPKMRGYQRLAKWIDGFAGALFAGF 198

QY 229 AIKIML 234
Db 199 GIHLII 204

Search completed: April 24, 2003, 18:22:42
Job time : 13 secs

```

GenCore version 5.1.4.p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:21:02 ; Search time 33 Seconds
(without alignments)
1473.549 Million cell updates/sec

Title: US-09-105-117K-2
Perfect score: 1191
Sequence: 1 MVMEIFITGLLGASLLS.....INVVAVVMTALAIKLMGM 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_nhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	864.5	72.6	231	2	Q8RQM4
2	349	29.3	202	16	Q8XXT9
3	347	29.1	211	16	Q8Z3W2
4	346	29.1	211	16	Q8ZM68
5	343	28.8	211	16	Q8XD10
6	336	28.2	212	16	Q8Y2F1
7	335	28.1	205	16	Q8ZHH6
8	332	27.9	200	16	Q8HW36
9	328.5	27.6	204	16	Q9X4K6
10	326.5	27.4	202	16	Q8UGV8
11	302.5	25.4	204	16	Q8Y87
12	285	23.9	226	16	Q9CKJ7
13	245.5	20.6	211	16	Q9KUN4
14	176.5	14.8	220	16	O06730
15	175.5	14.7	207	16	Q8RHX9
16	173.5	14.6	210	16	O25420

17	170.5	14.3	210	16	Q9ZLC4	Q9zlc4 helicobacte
18	164	13.8	205	16	Q99V11	Q99v11 staphylococ
19	154.5	13.0	205	16	Q97FY1	Q97fy1 clostridium
20	153.5	12.9	213	16	Q92HQ4	Q92hq4 rickettsia
21	151.5	12.7	208	17	Q27538	Q27538 methanobact
22	148	12.4	200	16	Q9KFP7	Q9kfp7 bacillus ha
23	136.5	11.5	206	16	Q8Y2B8	Q8y2b8 ralstonia s
24	136.5	11.5	213	16	Q913A2	Q913a2 pseudomonas
25	132	11.1	210	16	Q9AC37	Q9ac37 caulobacter
26	130	10.9	210	16	Q9K8S1	Q9k8s1 bacillus ha
27	129	10.8	210	16	Q983P3	Q983p3 rhizobium l
28	126	10.6	197	16	Q92NS0	Q92ns0 rhizobium m
29	120.5	10.1	208	16	Q9AAV6	Q9aav6 caulobacter
30	119	10.0	620	16	Q9KM59	Q9km59 vibrio chol
31	118	9.9	234	16	Q8UJW7	Q8ujw7 agrobacteri
32	117.5	9.9	197	16	Q9HZT1	Q9hzt1 pseudomonas
33	116	9.7	205	2	Q9RMX0	Q9rmx0 bacillus an
34	114.5	9.6	208	16	Q9K775	Q9k775 bacillus ha
35	113	9.5	222	16	Q9ABU4	Q9abu4 caulobacter
36	113	9.5	235	16	Q8UC29	Q8uc29 agrobacteri
37	112.5	9.4	211	16	Q92NK8	Q92nk8 rhizobium m
38	111.5	9.4	208	16	Q98G16	Q98g16 rhizobium l
39	111	9.3	158	2	Q9XBR8	Q9xbr8 zymomonas m
40	111	9.3	210	16	Q8Y176	Q8y176 ralstonia s
41	111	9.3	212	16	Q8XEY9	Q8xev9 salmonella
42	109.5	9.2	208	16	Q985J9	Q985j9 rhizobium l
43	109.5	9.2	211	16	Q92LY9	Q92ly9 rhizobium m
44	109.5	9.2	218	16	Q92U06	Q92u06 rhizobium m
45	109.5	9.2	450	17	Q9YC69	Q9yc69 aeropyrum p

ALIGNMENTS

RESULT 1

Q8RQM4 PRELIMINARY; PRT; 231 AA.
AC Q8RQM4;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Lysine exporter protein.
GN LYSE.
OS Corynebacterium efficiens.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RA Itaya H., Kimura E., Kawahara Y., Sugimoto S.;
RT "lyse, lyse of Corynebacterium efficiens.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB083133; BAB8827.1; -
SQ SEQUENCE 231 AA; 24776 MW; 8D14D4D7AD937E17 CRC64;

Query Match	72.6%	Score	864.5	DB 2	Length	231			
Best Local Similarity	71.2%	Pred. No.	9.3e-65						
Matches	166	Conservative	30	Mismatches	32	Indels	5	Gaps	2
QY	4	MEIFITGLLGASLLSICGPQNVLTQIKREGLIALLVCLISDVFLFIAGTIGVDLL	63						
Db	4	MEIFVTGLLGASLLAIGPQNVLTQIKREGLIALLVCLISDVFLFIAGTIGVDLL	63						
QY	64	SNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTKVPAQIIEETPTVPDDTFLGSSAVA	123						
Db	64	SDTAPIILDLRWCGIAYLLWFAVMAARDALRARTV-TFVHSEFPVAASAGSGG----	118						
QY	124	TDTNRVRVEVSDQKRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGAQYGDTCRWIFA	183						
Db	119	VTTKQRPRLRITSGTRQVWRPMLMAIVLTWLNPNAYLDADFVIGVGAQYGETGRWIFA	178						
QY	184	AGAFASLLWFFLVGFGAAALSRPLSSPKVWRWVNVVAVVMTALAIKLMWG	236						

Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.
RA Churser C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebailia M.,
BA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitlehead S., Barrrell B.G.;

"Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR ENBL; AL627277; CAD02896.1; --
DR InterPro; IPR001123; LySE.
DR InterPro; IPR004777; LyS_exporter.
DR Pfam; PF01810; LySB; 1.
DR TIGRFAMS; TIGR00948; 2a75; 1.
KW Complete proteome..
SQ SEQUENCE 211 AA; 23172 MW; 13CB427CABESA3FC CRC64;

Query Match 29.1%; Score 347; DB 16; Length 211;
Best Local Similarity 33.6%; Pred.No.1.8e-21;
Matches 79; Conservative 42; Mismatches 80; Indels 34; Gaps

Qy 3 IMEIFITGLLGASLLISGPONVLVYKGIKEGHLIAVLVCLLSIDVFELTAGTLGVDL 62
Db : ::::: |||:::||||:::||:::||:::||:::||:::||:::||:::
1 MISYFGVALGAAMILPLGPSQAFVMNQIRKHMTALCALSDLVLIISAGIFGGSA 60
Qy 63 LSNAAPTVLIDIMRWGGTAVLLNPAVMAAKDMTNKVKAPOIIEETPTVPDDTPPGGSV 122
Db 61 LLMQSPWLLALTVGWGVAFLNWGFGLATKMSSNLLE----- 98
Qy 123 ATDTRNRVRVEVSVDQRVVVKPMLMAIVLTWLNPNAVYLDAFVFIGVGAOYG-DTGRIWI 181
Db 99 -----ASAEBMKQGRW-KIIATWLAWTLNPHVYLDTFVVLGSLQGQLAKEPKRW- 147
Qy 182 FAAGAFAAASIWFPLGFGAAALSRLPSPKWRWNVNNAVVTALATKLMLMG 236
Db 148 FALGTTSASFNFEGALLANWLAPLRATAQRIINILGVNVVWLIALQOLAREG 202

RESULT 4
Q8ZM68 PRELIMINARY; PRT; 211 AA.

ID Q8ZM68 AC Q8ZM68 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative LYSE family, amino acid transport protein.
GN YGCA OR STM3066.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Salmonella.
OC NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lt2 / SGSC412 / ATCC 700720;
RK MDLINL21534948; PubMedId1677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grew N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
DR ENBL; AE008841; AAL21941.1; --
DR InterPro; IPR001123; LySE.
DR InterPro; IPR004777; LyS_exporter.
DR Pfam; PF01810; LySB; 1.
DR TIGRFAMS; TIGR00948; 2a75; 1.
SW Hypothetical protein; Complete proteome.
SQ SEQUENCE 211 AA; 23220 MW; 64B6FC8FF21FD0F CRC64;

Query Match 29.1%; Score 346; DB 16; Length 211;
 Best Local Similarity 33.6%; Pred. No. 2.2e-21;
 Matches 79; Conservative 41; Mismatches 81; Indels 34; Gaps 4;

QY 3 IMEIFTGGLGASLLISIGPQNVLVKQIKREGLIAVLVCLISDVFLFIAGTIGVDL 62
 DB 1 MISYFQGPALGAAMILPLGPQNAFVNMQGIROYHLMIALCALSDLVLSAGIFGSA 60
 QY 63 LSNAAPITVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAV 122
 DB 61 LLWQSPWLLALVTWGGVAFLLWTFGFKATMSSNLEL----- 98
 QY 123 ATDTRNRVREVSVDKQVRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQYG-DTGRWI 181
 DB 99 -----ASAEVLVQGRW-KIIATMLAVTWNPHVYLDTFVVLGSLGQLAMEPKRW- 147
 QY 182 FAAGAFAASLIWPLVPGFGAALSRLSPSSPKVWRWVNVVAVMTALAIKMLMG 236
 DB 148 FALGTISASPLWFFGLALLAALWAPRLRTAKAORIINILGVVVMWLIATFOLAREG 202

RESULT 5
 Q8XD10 PRELIMINARY; PRT; 211 AA.
 AC Q8XD10; (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
 DE Orf, hypothetical protein.
 GN YGGA OR Z4260 OR ECS3794.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]_TaxID=83334;
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takani H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22 (2001).
 DR EMBL; AE005522; AAG58049.1; -
 DR EMBL; AP002563; BAB37217.1; -
 DR InterPro; IPR001123; LysE.
 DR InterPro; IPR004777; Lys exporter.
 DR Pfam; PF01810; LysE; 1.
 DR TIGRFAMs; TIGR00948; 2a75; 1.
 KW Complete proteome.
 SQ SEQUENCE 211 AA; 23202 MW; 4422D078B182266C CRC64;

Query Match 28.8%; Score 343; DB 16; Length 211;
 Best Local Similarity 35.2%; Pred. No. 4e-21;
 Matches 80; Conservative 37; Mismatches 76; Indels 34; Gaps 4;

QY 7 FITGLLGASLLISIGPQNVLVKQIKREGLIAVLVCLISDVFLFIAGTIGVDL 66
 DB 5 YFOGLGAAMILPLGPQNAFVNMQGIROYHLMIALCALSDLVLSAGIFGSA 64

QY 67 APVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAVATDT 126
 DB 65 SPWLLALVTWGGVAFLLWTFGFKATMSSNLEL----- 98
 QY 127 RNRVREVSVDKQVRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQYG-DTGRWIFAA 185
 DB 99 -----ASAEVLVQGRW-KIIATMLAVTWNPHVYLDTFVVLGSLGQLDVEPKRW-FALG 151
 QY 186 AFAASLIWPLVPGFGAALSRLSPSSPKVWRWVNVVAVMTALAIKL 232
 DB 152 TISASPLWFFGLALLAALWAPRLRTAKSORIINILVGVVCMWFIALQL 198

RESULT 6

Q8Y2F1 PRELIMINARY; PRT; 212 AA.
 AC Q8Y2F1; (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
 DE Probable transmembrane protein.
 GN RSC0385 OR RSO3353.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]_TaxID=305;
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunin S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502 (2002).
 DR EMBL; AL646059; CAD13913.1; -
 DR InterPro; IPR001123; LysE.
 DR Pfam; PF01810; LysE; 1.
 KW Complete proteome.
 SQ SEQUENCE 212 AA; 21896 MW; E29D663C9497EDBB CRC64;

Query Match 28.2%; Score 336; DB 16; Length 212;
 Best Local Similarity 31.0%; Pred. No. 1.5e-20;
 Matches 71; Conservative 44; Mismatches 82; Indels 32; Gaps 1;

QY 8 ITGLLGASLLISIGPQNVLVKQIKREGLIAVLVCLISDVFLFIAGTIGVDL 67
 DB 16 LSGFGLGASLVAIGAQAQYVLRQGRREYVLRVGLVLCALCDMALIALGVAGMGTLSAH 75
 QY 68 FIVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAVATDT 127
 DB 76 PAWLTAVERWAGAAFLAYGARAFRAAW-----R 103
 QY 128 RNRVREVSVDKQVRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQYDGTGRWIFAA 187
 DB 104 GAERLQARNGDKASHAQVLSALALSLLNPHVYLDTFVVLGALIGGRYAMPANPAPAGAM 163
 QY 188 AASLIWPLVPGFGAALSRLSPSSPKVWRWVNVVAVMTALAIKMLMG 236
 DB 164 CASILWFLSLGFGARLLPEVPFVPMRVLDALIGAVNMWALITLLMGG 212

RESULT 7

Q8ZHH6 PRELIMINARY; PRT; 205 AA.
 AC Q8ZHH6; (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)

DE Putative LysE type translocator.
GN YPO0918.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21740413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feitwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ414145; CAC89762.1; -;
DR InterPro; IPR001123; LysE.
DR InterPro; IPR004777; Lys_exporter.
DR Pfam; PF01810; LysE; 1.
DR TIGRFAMs; TIGR00948; 2a75; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 22164 MW; 2969524391474 CDA CRC64;
Query Match 28.1%; Score 335; DB 16; Length 205;
Best Local Similarity 33.1%; Pred. NO. 1.8e-20;
Matches 80; Conservative 44; Mismatches 70; Indels 48; Gaps 6;
QY 3 IMEFTFGLGASLLSISGPONVIVIKQIKREGIAVLVCLISDVFLFIAGTIGVDL 62
DB 1 MLAVYLHGFLSAAILPLGPQNVVNMQIKRQHLMSSALCALSDIILICAGIFGSA 60
QY 63 LSNAAPIVLDIMRWGGIAYLLWF---AVMAAKDAMTKNVEAPQIIETEPTVDDTPLGG 119
DB 61 LLSRSPILLALVYGGVAFPLMWYWGALMA-----WRC 94
QY 120 SAVATD---TRNRVREVSVDQRVWVKPMLMAIVLTWLNPNAYILDFAVFIGVGQAQ-Y 174
DB 95 DGVASSATSVTQGRWRLVPL-----LAVTLNPHVYLDTFVVLGSLGQLL 141
QY 175 GDTGRWIFACAFAPASILVPLVGFAGAAALSRPLSSPKVWRWVNVVAVMTALAKML 234
DB 142 PDIRPW-FALGAVTASIVWFFALALLAALWLSPLNRPVQAQRIINLFGVGMGFIAFQAR 200
QY 235 MG 236
DB 201 QG 202
RESULT 8
Q9HW36 PRELIMINARY; PRT; 200 AA.
ID Q9HW36
AC Q9HW36;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Probable transporter.
GN PA4365.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey R.K., Bramkan P.S.L., Huynh W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AE004852; AAG07753.1; -;
DR InterPro; IPR002109; Glutaredoxin.
DR InterPro; IPR001123; LysE.
DR Pfam; PF01810; LysE; 1.
DR PROSITE; PS00195; GLUTAREDOXIN; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 200 AA; 21144 MW; 7C10CB8D8FE36A4 CRC64;
Query Match 27.9%; Score 332; DB 16; Length 200;
Best Local Similarity 33.5%; Pred. NO. 3.1e-20;
Matches 77; Conservative 42; Mismatches 77; Indels 34; Gaps 4;
QY 7 FITGLLGASLLSISGPONVIVIKQIKREGIAVLVCLISDVFLFIAGTIGVDLSNA 66
DB 5 YLNGILVAAGLIIAIGAQNAFVLAQSLRHHLSVAALCVFCDAVLVSLGVFLAKLLE 64
QY 67 APIVLDIMRWGGIAYLLWFVAVMAAKDAMTKNVEAPQIIETEPTVDDTPLGGSAVATDT 126
DB 65 NPTLLAIARWGGIAYFLTWYGLKALLR-----PDAL--GNAETGP 105
QY 127 RNRVREVSVDQRVWVKPMLMAIVLTWLNPNAYILDFAVFIGVGQAQYDGTGRWIFAA 186
DB 106 RSRKAV-----LLAALAVTLNPHVYLDTFVVLGSLGQAQAPG--AYALGA 150
QY 187 FASLIWPLVPLVGFAGAAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 236
DB 151 ASASLWFFPALGAALWAPLARPATWRLDLVAVAMMLGWAQQLFRG 200
RESULT 9
Q9K4K6 PRELIMINARY; PRT; 204 AA.
ID Q9K4K6
AC Q9K4K6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative membrane transport protein.
GN SC07308 OR SC5F8.18.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmid and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
RT *cholerae*.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004134; AAF93654.1; -.
DR TIGR; VC0481; -.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
DR Complete proteome.
SQ SEQUENCE 211 AA; 22651 MW; C92B752CCB6B0741 CRC64;

Query Match 20.6%; Score 245.5; DB 16; Length 211;
Best Local Similarity 28.6%; Pred. No. 5.8e-13;
Matches 65; Conservative 34; Mismatches 97; Indels 31; Gaps

QY 6 IFITGLLLGASLLLSIGFONVIVKQIKREGLIAVLVLCILSDVFIFAGTIGVDLLSN 65
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 5 ILLQFSLGATMIPIGAQNAYVNLQIKRHHLTLTAATCGVLDMIFITLIGFGGALIS 64
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 66 AAPVILDMWGGIAYLLFAVMAAKDAMTKNKEAPQIIETETVPDDTFLGSSAVATD 125
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 65 QNTSLLTGVTLAGTILFCGGVLSRAAL-----KPPQSESTA-----NPMAGRKA- 113
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 126 TRNRVRVEVSVDQRVMVKMLMAVLTLNPNAYLDADFVFIGVGQAQYGDTRGWIFAAG 188
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 114 -----IFGAFATVNPHLYLDTTVILSIGQGQFQDERISPAG 153
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 186 AFASLIWFPLVGFGAALSRLPSPKVWRWINVVAVMVALAIKL 232
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 154 TILASFVWFFTLSIGAALKSTLTSLKPRVRVIDMAVAAMFFIAFAL 200
| : ||||:::||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 14
O06730 PRELIMINARY; PRT; 220 AA.

ID OC O06730; SEQUENCE FROM N.A.
AC O06730; MEDLINE=98015416; PubMed=9353932;
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JAN-1998 (TEMBLrel. 05, last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, last annotation update)
DE YISU protein.
GN YISU.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98015416; PubMed=9353932;
RA Roche B., Autret S., Levine A., Vannier F., Medina N., Seror S.J.;
RT "A *Bacillus subtilis* chromosome segment at the 100 degrees to 102
RT degrees position encoding 11 membrane proteins."
RL Microbiology 143:3309-3312(1997).
RC [2]
PC SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Avevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devigne K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gim S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haitech J., Harwood C.R., Henaut A.,
RA Horlbert H., Holsappel S., Hosono S., Hullio M.P., Itaya M., Jones L.,
RA Joris B., Karanata D., Kasahara Y., Klaerr-Bianchand M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Mael C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weizenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis.";
 RL Nature 390:249-256 (1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; Y09476; CAA70650.1; -;
 DR EMBL; 299109; CAB12926.1; -;
 DR InterPro; IPR001123; Lyse.
 DR InterPro; IPR004777; Lys_exporter.
 DR Pfam; PF01810; Lyse; 1.
 DR TIGRFAMs; TIGR00948; 2a75; 1.
 KW Complete proteome.
 SQ SEQUENCE 220 AA; 24065 MW; C18C2810F44F9656 CRC64;

Query Match 14.8%; Score 176.5; DB 16; Length 220;
 Best Local Similarity 22.9%; Pred. No. 3.6e-07;
 Matches 53; Conservative 36; Mismatches 99; Indels 43; Gaps 4;
 QY 4 MEIFITGLLGASLLLSIGPQNVLVKQIKREGL---TAVLVCLISDVFLFIAGTLGV 60
 DB 18 MNAIHGIVLAFGLILPLGVQNVFIQOQALOKHWRALPAVISASVCDTLILVLAVAGV 77
 QY 61 DLLSNAAPIVLDIRWGGIAYLLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGS 120
 DB 78 SVIVQELPVPETVMAGGFLFLYMGV-----105
 QY 121 AVATDTRNRVRVSVVDKQVWV--RPLMAI VLTWLNPNAYLDFAVFVIGGVGAOYGDG 178
 DB 106 -----TWN-IRPNTSQNKHTFTPKQAFAAFAAASVLLNPHAILDTIGVIGTSLSQSGLE 159
 QY 179 RMTFAGAGFAAASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALA 229
 DB 160 KWLFAACIAVSWIWFISLAIGRLFTIDTSGRLMLIVNKCSAAVWMAA 210

RESULT 15
 Q8RH9 PRELIMINARY; PRT; 207 AA.
 AC Q8RH9;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE L-lysine permease.
 GN FN1861.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
 RA Fonstein M., Kyrpides N., Overbeek R.;

RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018 (2002).
 DR EMBL; AE010488; AAL93960.1; -;
 KW Complete proteome.
 SQ SEQUENCE 207 AA; 23192 MW; 5833132210F1A771 CRC64;
 Query Match 14.7%; Score 175.5; DB 16; Length 207;
 Best Local Similarity 24.5%; Pred. No. 4.1e-07;
 Matches 58; Conservative 35; Mismatches 97; Indels 47; Gaps 5;
 QY 4 MEIFITGLLGASLLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGYD 61
 DB 1 MEKYLQGLMGLAYVAPIGVQNLVINSAITQKRSKALLIALIVIFEDVTLAFACFFGIG 60
 QY 62 DLLSNAAPIVLDIRWGGIAYLLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGSSA 121
 DB 61 LL-----IDKLEWLKLIIL-----VGSIV 80
 QY 122 VATDTRNRVRVSVVDKQVWVVKPMLMAI---VLTWLNPNAYLDFAVFVIGGVGAOYGD 177
 DB 81 IYIGQGLRSKSELKNDNDMDIPILKAITSCVVTWFPQAIIDGTWMLGAFRATLSSE 140
 QY 178 GRWIFAGAGFAAASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALA 234
 DB 141 AGIYFILGVTSASFVCFWMLSLISFISLFSHKFND-KVLKVINVCGLVITFYGVKLL 196

Search completed: April 24, 2003, 18:23:23
 Job time : 36 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:20:02 ; Search time 36 Seconds
(without alignments)
873.532 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVMETFTCLLGASLLLS.....INVVAVVMTALAKLMLMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1191	100.0	236	18 AAW37715	C. glutamicum Lys
2	1191	100.0	236	22 AAG93201	C. glutamicum prote
3	1178	98.9	233	22 AAU71888	C. glutamicum meta
4	1178	98.9	233	22 AAB79660	Corynebacterium gl
5	864.5	72.6	231	22 AAG64047	Corynebacterium th
6	350	29.4	211	21 AAB01789	Escherichia coli Y
7	170.5	14.3	210	18 AAW20426	H. pylori surface
8	170.5	14.3	215	18 AAW20968	H. pylori transmem
9	132.5	11.1	226	22 AAG89911	C. glutamicum prote
10	119.5	10.0	173	23 ABP39181	Staphylococcus epi

11	114.5	9.6	229	21 AAB11637	A. vitis hypersens
12	105	8.8	206	21 AAY99598	E. coli L-threonin
13	101	8.5	212	21 AAB01787	Escherichia coli Y
14	95	8.0	390	22 AAU58229	Propionibacterium
15	93.5	7.9	452	22 AAU34590	E. coli cellular p
16	93	7.8	459	22 AAG93244	C. glutamicum prote
17	93	7.8	459	22 AAB78987	C. glutamicum SRT
18	92	7.7	223	21 AAB01786	Escherichia coli Y
19	92	7.7	694	23 ABB93711	Herbicidally activ
20	91.5	7.7	446	22 AAG92602	C. glutamicum prote
21	91.5	7.7	446	22 AAB76722	Corynebacterium gl
22	91.5	7.7	446	23 AAE22703	Corynebacterium gl
23	91	7.6	111	20 AAW89961	Antigen from clust
24	91	7.6	624	22 AAG90535	C. glutamicum prote
25	91	7.6	628	22 AAB76784	Corynebacterium gl
26	91	7.6	628	22 AAB76842	Corynebacterium gl
27	90	7.6	4572	19 AAW52845	A. mediterranei ri
28	89.5	7.5	205	21 AAY79298	E. coli L-homoseri
29	89.5	7.5	205	21 AAY79298	E. coli RhtB prote
30	89.5	7.5	340	22 AAU36387	Pseudomonas aerugi
31	89	7.5	123	22 ABG23207	Novel human diagno
32	89	7.5	716	22 AAU64250	Propionibacterium
33	88	7.4	406	22 AAU39609	Propionibacterium
34	88	7.4	1308	22 ABB62870	Drosophila melanog
35	86.5	7.3	195	21 AAB01788	Escherichia coli Y
36	86.5	7.3	279	22 AAU55142	Propionibacterium
37	86	7.2	295	22 AAU34511	E. coli cellular p
38	86	7.2	400	22 AAG90734	C. glutamicum prote
39	85.5	7.2	138	22 AAG82080	S. epidermidis ope
40	85.5	7.2	224	23 ABP39763	Staphylococcus epi
41	85	7.1	548	22 ABB62090	Drosophila melanog
42	84.5	7.1	415	23 AAO22164	Ramoplanin biosynt
43	84.5	7.1	810	23 ABB93675	Herbicidally activ
44	84	7.1	514	23 ABB91817	Herbicidally activ
45	83.5	7.0	210	22 AAG81807	S. epidermidis ope

ALIGNMENTS

RESULT 1
AAW37715
ID AAW37715 standard; Protein; 236 AA.
AC AAW37715;
XX
DT 12-MAR-1998 (first entry)
XX
DE C. glutamicum Lys E protein (lysine export protein).
XX
KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
KW Microbial production; amino acid; animal feed additive.
XX
OS Corynebacterium glutamicum.
XX
PN DE19548222-A1.
XX
PD 26-JUN-1997.
XX
PF 22-DEC-1995; 95DE-1048222.
XX
PR 22-DEC-1995; 95DE-1048222.
XX
(KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
PA Eggeling L, Sahn H, Vrljic M;
PI
XX
DR WPI; 1997-333867/31.
DR N-PSDB; AAT96816.
XX
PT Increasing microbial production of amino acids, especially lysine -
PT by improving export carrier activity or corresponding gene
PT expression, also new export and regulatory genes from

PT Corynebacterium
XX
PS Claim 42; Page 10; 16pp; German.
XX

CC This sequence is the LysE protein product, a lysine export protein.
CC LysG and LysB encode a lysine transport regulatory protein and an export
CC protein, respectively. Microbial production of amino acids (A) is
CC improved by increasing the export-carrier activity and/or the export gene
CC expression in a microorganism that produces (A). The method is
CC specifically used to increase production of lysine, used as an animal
CC feed additive. Other (A) are variously useful as pharmaceuticals,
CC condiments and intermediates for fine chemicals. This method increases
CC the amount of (A) secreted into the culture medium. Export of (A) has
CC been found to depend on a single gene.
XX

SQ Sequence 236 AA;
Query Match 100.0%; Score 1191; DB 18; Length 236;
Best Local Similarity 100.0%; Pred. No. 2e-128;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIMEIFITGLLGASLLSITGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60
DB 1 MVIMEIFITGLLGASLLSITGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60
QY 61 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGS 120
DB 61 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGS 120
QY 121 AVATDTRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDADFVFIGGVAQYGDGTGRW 180
DB 121 AVATDTRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDADFVFIGGVAQYGDGTGRW 180
QY 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236
DB 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236

RESULT 2
AAG93201
ID AAG93201 standard; Protein; 236 AA.
XX
AC AAG93201;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ. ID NO: 6955.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
XX
XX EPI108790-A2.
XX
XX 20-JUN-2001.
XX
XX 18-DEC-2000; 2000EP-0127698.
XX
XX 16-DEC-1999; 99JP-0377484.
XX 07-APR-2000; 2000JP-0159162.
XX 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
XX N₂PSDB; AAH68420.
XX
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
XX mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 29; SEQ ID NO: 6955; 246pp + Sequence Listing; English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or identifying a homologue of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX

SQ Sequence 236 AA;
Query Match 100.0%; Score 1191; DB 22; Length 236;
Best Local Similarity 100.0%; Pred. No. 2e-128;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIMEIFITGLLGASLLSITGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60
DB 1 MVIMEIFITGLLGASLLSITGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60
QY 61 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGS 120
DB 61 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGS 120
QY 121 AVATDTRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDADFVFIGGVAQYGDGTGRW 180
DB 121 AVATDTRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDADFVFIGGVAQYGDGTGRW 180
QY 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236
DB 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236

RESULT 3
AAU71888
ID AAU71888 standard; Protein; 233 AA.
XX
AC AAU71888;
XX
DT 26-FEB-2002 (first entry)
XX
DE C. glutamicum metabolic pathway protein encoded by gene #23.
XX
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
XX methionine biosynthesis pathway; large-scale production of fine chemical;
XX Corynebacterium diphtheriae; diphtheria.
XX
XX Corynebacterium glutamicum.
XX
XX WO200166573-A2.
XX
XX 13-SEP-2001.
XX
XX 22-DEC-2000; 2000WO-IB02035.
XX
XX 09-MAR-2000; 2000US-187970P.
XX 23-JUN-2000; 2000US-0606740.
XX
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habehauer G, Kim J;
XX Lee H, Hwang B;
XX WPI; 2001-582269/65.
XX

DR N-PSDB; AAS96099.
 XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
 PT glutamicum, useful for producing methionine and lysine in
 PT Corynebacterium and Brevibacterium -
 XX
 PS Disclosure; Page 215-216; 316pp; English.
 XX
 CC The present invention relates to the isolation of novel Corynebacterium
 CC glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)
 CC proteins. The metabolic pathway proteins of the invention
 CC include enzymes involved in the lysine and methionine biosynthetic
 CC pathways. The polynucleotide sequences of the invention can be used
 CC for the large-scale production and/or modulation of expression of the
 CC fine chemicals such as lysine and methionine. The sequences of the
 CC invention may be used to identify C. glutamicum and related organisms
 CC e.g. C. diphtheriae in a subject to detect diphtheria.
 CC AAU71863-AAU71922 represent the novel C. glutamicum metabolic pathway
 CC proteins of the invention.
 XX
 SQ Sequence 233 AA;
 Query Match 98.9%; Score 1178; DB 22; Length 233;
 Best Local Similarity 100.0%; Pred. No. 6.1e-127;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 MEIFITGLGASLLLSIGFQNVLVKQIKREGILAVLLVCLISDVFLFIAGTLGVDLL 63
 DB 1 MEIFITGLGASLLLSIGFQNVLVKQIKREGILAVLLVCLISDVFLFIAGTLGVDLL 60
 QY 64 SNAPIVLDIMRWGGIAYLWFAVMAAKDAMTKVEAPQIIETEPTVPDDTLGGSAVA 123
 DB 61 SNAPIVLDIMRWGGIAYLWFAVMAAKDAMTKVEAPQIIETEPTVPDDTLGGSAVA 120
 QY 124 TDTNRVRVSVSKQVWVKPMLMAVLTLWLNPNAYLDAFVFIGGGAQYGTGRWIFA 183
 DB 121 TDTNRVRVSVSKQVWVKPMLMAVLTLWLNPNAYLDAFVFIGGGAQYGTGRWIFA 180
 QY 184 AGAFAASLIWFLVGFCAALSRPLSPKVRWVNVVAVVMTALAKLMG 236
 DB 181 AGAFAASLIWFLVGFCAALSRPLSPKVRWVNVVAVVMTALAKLMG 233
 RESULT 4
 AAB79660
 ID AAB79660 standard; Protein; 233 AA.
 AC AAB79660;
 XX
 XX 30-APR-2001 (first entry)
 DT
 DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:54.
 XX
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00923.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032130.
 PR 09-JUL-1999; 99DE-1032186.
 PR 09-JUL-1999; 99DE-1032206.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.

(BADI) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI; 2001-137957/14.

N-PSDB; AAF71779.

XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -

Claim 20; Page 234-235; 1737pp; English.

XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic

CC compounds, vitamins, cofactors, polyketides and enzymes.

XX
SQ Sequence 233 AA;
Query Match 98.9%; Score 1178; DB 22; Length 233;
Best Local Similarity 100.0%; Pred. No. 6.1e-127;
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 63
DB 1 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 60

QY 64 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 123
DB 61 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 120

QY 124 TDTNRVRVSVSKQVWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 183
DB 121 TDTNRVRVSVSKQVWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 180

QY 184 AGAPAAASLIWFLVGFAGAAALSRLSPKVRWVWVAVVMTALAIAKMLMG 236
DB 181 AGAPAAASLIWFLVGFAGAAALSRLSPKVRWVWVAVVMTALAIAKMLMG 233

RESULT 5
AAG64047
ID AAG64047 standard; Protein; 231 AA.
AC AAG64047;
XX
DT 11-SEP-2001 (first entry)
XX
DE Corynebacterium thermoaminogenes lysin biosynthetic enzyme lyse.
XX
KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;
KW aspartate-semialdehyde dehydrogenase; lyse.
XX
OS Corynebacterium thermoaminogenes.
XX
PN JP2001120270-A.
XX
PD 08-MAY-2001.
XX
PF 01-NOV-1999; 99JP-0311148.
XX
PR 01-NOV-1999; 99JP-0311148.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI; 2001-364760/38.
DR N-PSDB; AAA45375.
XX
XX
XX A heat-resistant lysin biosynthetic system enzyme gene of a high temperature-resistant coryneform microbe - Japanese.
XX
PS Claim 7; Page 24; 27pp; Japanese.
XX
CC The invention relates to a gene from a high temperature-resistant coryneform microbe that encodes a heat-resistant lysin biosynthetic enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity and can be used for growing amino acid-producing microbes. The present amino acid sequence corresponds to an enzyme of the invention.
XX
SQ Sequence 231 AA;
Query Match 72.6%; Score 864.5; DB 22; Length 231;
Best Local Similarity 71.2%; Pred. No. 6.7e-91;
Matches 166; Conservative 30; Mismatches 32; Indels 5; Gaps 2;

QY 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 63
DB 1 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 60

DB 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 63
QY 64 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 123
DB 64 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 118
QY 124 TDTNRVRVSVSKQVWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 183
DB 119 VTTKQPRLRITSGTQVWVRPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 178
QY 184 AGAPAAASLIWFLVGFAGAAALSRLSPKVRWVWVAVVMTALAIAKMLMG 236
DB 179 AGAPAAASLIWFLVGFAGAAALSRLSPKVRWVWVAVVMTALAIAKMLMG 231

RESULT 6
AAB01789
ID AAB01789 standard; Protein; 211 AA.
XX
AC AAB01789;
XX
DT 03-JAN-2001 (first entry)
XX
DE Escherichia coli YggA amino acid excretion protein.
XX
KW E. coli; YggA gene; amino acid production; excretion protein gene;
KW amino acid excretion protein.
XX
OS Escherichia coli.
XX
PN EP1016710-A2.
XX
PD 05-JUL-2000.
XX
PF 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1999; 98RU-0124016.
PR 09-MAR-1999; 98RU-0104431.
XX
PA (AJIN) AJINOMOTO CO INC.
XX
PI Livshits VA, Zakataeva NP, Nakanishi K, Alekhin VV, Troshin PV;
PI Tokhmakova IL;
XX
DR WPI; 2000-414802/36.
DR N-PSDB; AAA52691.
XX
PT Increased production of L-amino acids by an Escherichia bacterium comprises increasing the expression amount of an L-amino acid excretion protein -
XX
PS Claim 1; Page 25; 29pp; English.
XX
CC The present sequence is the YggA amino acid excretion protein from Escherichia coli. This protein is involved in the production of amino acids, and an increase in its expression leads to an increased accumulation of amino acids in the cell. In this case, an increase in arginine, glutamic acid and lysine is achieved if multiple copies of the gene are transfected into a bacterium. The bacterium used is E. coli.
XX
SQ Sequence 211 AA;
Query Match 29.4%; Score 350; DB 21; Length 211;
Best Local Similarity 35.7%; Pred. No. 8.4e-32;
Matches 81; Conservative 37; Mismatches 75; Indels 34; Gaps 4;

QY 7 FITGALGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLLSNA 66
DB 5 YFQGLAGAAITLPGPQVAFVWVQVIRYHIMTIALCAISDLVLCAGIFGGSALLMQ 64
QY 67 APTVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 126
DB 65 SPWLLALVTWGGVAFLLWYGFAGFTAMSSNIEL----- 98

QY 127 RNRVRVEVSDKQVWVKPMLMAVLTLNPNAYLDAFVFIGGVAQV-DTGRWIFAA 185
 Db 99 -----ASAENVKQGRW-KIATMLAVTLNPHVYLDTFVVLGSLGGQDVEPKW-FALG 151
 QY 186 APAASLIWFPLVGGAAALSRPLSSPKVRWVWVAVVMTALAIAKL 232
 Db 152 TISASFLWFFGLALLAAWAPRLRTAKAQRITINLVGCVWVFIALQL 198

RESULT 7
 AAW20426
 ID AAW20426 standard; Protein; 210 AA.
 XX AAW20426;
 XX
 DT 14-JUL-1997 (first entry)
 XX
 DE H. pylori surface membrane protein 33986087.aa.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 WPI: 1997-052306/05.
 DR N-PSDB; AAT67601.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 604; 1481pp; English.

CC The present sequence shows a Helicobacter pylori surface membrane
 CC protein, having five transmembrane regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 210 AA;
 Query Match 14.3%; Score 170.5; DB 18; Length 210;
 Best Local Similarity 21.5%; Pred. No. 3.6e-11;
 Matches 50; Conservative 48; Mismatches 92; Indels 43; Gaps 5;

QY 6 IFITGLLGSALLSTGPNVIVIKQIKREGIAVLVCLISDVFLTAGTGVDLN 65
 Db 4 VFIEGFLAISLCAAVGAQSLFIVERGARNYVFLICALCFMCDIVLMSGVGVGAYF- 62

QY 66 AAPIVLDI-MRWGGIAYLLWFVAWMAKDAWTKNKEAPQIIIEETPTVPDDTPIGGSAVAT 124
 Db 63 AKNLYLSLFINFGAVFTGTGYAFLAK-----T 90
 QY 125 DTRNRVRVEVSDKQVWVKPMLMAVLTLNPNAYLDAFVFIGGVAQV-DTGRWIFAA 184
 Db 91 LFGTFKKKQVQTPKKLSLKKTLTLFTLGVTLNPNQVYLEWVFLIGASAMSFNLVQKVFLLA 150
 QY 185 GAFAASLIWFPL-----VGFGAALSRPLSSPKVRWVWVAVVMTALAIAKL 232
 Db 151 GTLSAASFSLLLCTMSLYG-----SKLNNQKIPMGVNLFTVTAINGTSLVTL 199

RESULT 8
 AAW20968
 ID AAW20968 standard; Protein; 215 AA.
 XX AAW20968;
 XX
 DT 21-JUL-1997 (first entry)
 XX
 DE H. pylori transmembrane protein, hp2p10625orf14.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PN WO9640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaard BL;
 XX
 WPI: 1997-052306/05.
 DR N-PSDB; AAT68221.
 XX

PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 73; Page 1359; 1481pp; English.
 XX
 CC This sequence represents a H. pylori protein likely to contain five
 CC membrane spanning regions.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX

SQ Sequence 215 AA;
 Query Match 14.3%; Score 170.5; DB 18; Length 215;
 Best Local Similarity 21.5%; Pred. No. 3.7e-11;

SQ	Sequence	226 AA;
Query Match	11.1%; Score 132.5; DB 22; Length 226;	
Best Local Similarity	24.2%; Pred. No. 9.3e-07;	
Matches	59; Conservative 45; Mismatches 98; Indels 39; Gaps 12;	
QY .	7 FITCLLGASLLISIGPQNVLVIKQGIK--REGLI AVLVLVCLISDVFLFIAGTLCVDLLS 64 : : : : : : : : : : : : : : : : : : : : : : : : : : : : Db 6 FLALFLVMIAAISPGLDFQIIRLSAKNRDGVLTAVGIMVGNSIW-I-TASLLGLSALI 64 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	65 NAAPITVIDMRWGCIAYLLWFANMAADAMTNKVPAQIIETEETPTVPDDTPPLGGSAVAT 124 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	65 STYPAILNLQLVGGGYLTWVGIGAVESWTKR-----STQAAADSQAVENTLVTA 116 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	125 DTRNRRVEVSVDQRVVWXPMLMAIVLTM-LNPNAVILDAPVFVIGGVGAQY--GDTC-RW 180 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	117 TAAS-----VCVMPAIRSGIATNLSNKA VL----FFGSVFAQFVRPDMGIGW 160 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
QY	181 IFRAGAF--AASLWFPVLGFGAALSRPLSS--PKVWRMINVVAVUMTALA IKLMLMG 236 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	161 SIFIGVFELTLGLWLF--VGF--AVLRKLAAGLTRNGAIIDLTTGVTFIGLGMFMIFEG 216 : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
RESULT 10		
ABP39181	ID ABP39181 standard; Protein; 173 AA.	
XX AC	ABP39181;	
XX DT	24-JUL-2002 (first entry)	
XX DE	Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4026.	
XX KW	Staphylococcus epidermidis; open reading frame; ORF; bacterial infection; antibacterial; gene therapy.	
XX OS	Staphylococcus epidermidis.	
XX PN	US6380370-B1.	
XX PD	30-APR-2002.	
XX PF	13-AUG-1998; 98US-0134001.	
XX PR	14-AUG-1997; 97US-055779P.	
XX PR	08-NOV-1997; 97US-064964P.	
XX PA	(GENO-) GENOME THERAPEUTICS CORP.	
XX PI	Doucette-Stamm LA, Bush D;	
XX DR	WPI; 2002-381255/41.	
XX DR	N-PSDB; ABN91726.	
XX FT	Novel isolated nucleic acid encoding a Staphylococcus epidermidis polypeptide, useful for diagnosing and treating bacterial infections -	
XX PS	Disclosure; SEQ ID 4026; 267pp; English.	
XX CC	ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading frame (ORF) nucleic acid sequences which encode the amino acid sequences given in ABP35124 to ABP37960. The S. epidermidis sequences have antibacterial activity and can be used in gene therapy. The sequences can also be used in the diagnosis and treatment of bacterial infections, particularly S. epidermidis infections. The sequences can be used to screen for compounds able to interfere with the S. epidermidis life cycle or inhibit S. epidermidis infection.	
XX CC	N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the USPTO web site.	
SQ	Sequence	173 AA;

Query Match 10.0%; Score 119.5; DB 23; Length 173;
Best Local Similarity 23.0%; Pred. No. 2e-05;
Matches 46; Conservative 32; Mismatches 83; Indels 39; Gaps 4

QY 39 IAVLLVCLISDFLEFTAGTLGVLDLILSNAAPIVLDIRMGGLIAYILLFAFVAAMAKDAMTNKV 98
 |::| | | | | | | | | | | | | | | | : | | : | |:
Db 8 LPWIIITAGLCDTELIVAILGSLHLSIPTQLFIYIIGFLFYLM-----ANSWLT 60
 |::| | | | | | | | | | | | | | | | : | | : | |:
QY 99 EAPQIIEETEPTVPDDTPLGGSAVATDTNRVRVEVSVDKQRVMVKPMLMAIVLTWLNP 158
 | | | | | | | | | | | | | | | | : | | : | |:
Db 61 EKPSNTEETEP-----MSAKQ-----ILFALSVELLNPH 90
 | | | | | | | | | | | | | | | | : | | : | |:
QY 159 AYLDAFPFGCGCAQYGDTGRWFPAAGAFAASLIWPPLVGFGAALSRPLSPKVKRWIN 218
 | | | | | | | | | | | | | | | | : | | : | |:
Db 91 AIMDTGVGIATSASVDGYDKVFSLATISSVIFWFLAILGRITKIDSKGYIVILN 150
 | | | | | | | | | | | | | | | | : | | : | |:

QY 219 VW--VAVMTALAIKLMLMG 236
 | | | | | | | | | | | | | | | | : | | : | |:
Db 151 KVVSVIVIIVGLIILKNIVG 170
 | | | | | | | | | | | | | | | | : | | : | |:

RESULT 11
AAB11637
ID AAB11637 standard; Protein; 229 AA.
XX
AC AAB11637;
XX
DT 23-OCT-2000 (first entry)
XX
DE A. vitis hypersensitive response elicitor protein, SEQ ID NO:9.
XX
KW Hypersensitive response elicitor protein; HR; disease resistance;
insecticide; fungicide; antiviral; bactericide; growth enhancer;
stress resistance; transgenic plant.
XW Agrobacterium vitis.
OS
XX
FN W0200028056-A2.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26079.
XX
PR 06-NOV-1998; 98US-0107387.
XX
PP (CORR) CORNELL RES FOUND INC.
XA
PA Burr TJ, Herlache TC, Zhang H;
XX
PI WPI; 2000-376567/32.
DR N-PSDB; AAA61502.
XX
XX New protein from Agrobacterium vitis, useful e.g. for imparting
resistance to disease or stress to plants, is involved in production of
a hypersensitive response -
XX
XX Claim 3; Page 85-86; 157pp; English.

Sequences AAB11630-B11688 represent proteins from Agrobacterium vitis which elicit a hypersensitivity response (HR) in a plant. The invention also relates to nucleotide sequences (AAA61501-A61524) encoding the A. vitis HR elicitor proteins. The HR is a rapid, localised necrosis that is associated with the active defence of plants against many pathogens, and occurs when a pathogenic organism interacts with a nonhost plant (i.e. one in which intracellular bacterial growth and disease development do not occur). Like other HR elicitors, the A. vitis elicitor functions in non-host plants by causing a rapid hypersensitive response that results in walling-off and killing of the pathogen. On grape plants, the A. vitis elicitor induces a restricted necrosis of tissues, resulting in the death of plant cells and induction of pathogen resistance. A. vitis HR elicitor proteins, in non-infectious form, are used to treat plants or their seeds to impart resistance to disease, such as those caused by fungi, bacteria or viruses; and to enhance growth, e.g., to increase

CC The present sequence is the L-threonine resistance protein, RhtB, from
 CC Escherichia coli. The coding sequence may be used to impart L-threonine
 CC resistance on E. coli bacteria, which would be useful for producing
 CC a high yield of L-threonine. L-threonine resistance means that the
 CC bacteria will be able to grow on a minimal medium containing
 CC L-threonine at a concentration at which the corresponding wild-type
 CC strain would not grow. Since the transformed bacteria can grow on the
 CC minimal medium, it can synthesize L-threonine, which accumulates. The
 CC accumulated amino acids can then be removed from the culture medium.
 CC The bacterium of the present invention may also be used to synthesize
 CC L-homoserine, L-valine and L-leucine at increased levels.

XX SQ Sequence 206 AA;

Query Match 8.8%; Score 105; DB 21; Length 206;

Best Local Similarity 20.4%; Pred. No. 0.0012;

Matches 51; Conservative 36; Mismatches 99; Indels 64; Gaps 9;

QY 3 IMEFTITGLLGLSLLSIGPQNVLVTKQIKR---EGLTAVLLVCLISDVFLFIAGTIG 59

DB 1 MLMLFLTVAMVHIVALMSPGDFFFVSQTAVSRKEMMGVLGITCGVMVWAGIA-LLG 59

QY 60 VDLISNAPIVLDIMRW-----GGIAYLLWFAVMAAKDAMTK---VEAPQIIEETEP 109

DB 60 LHL-----LIEKMAWHLTLMVGGGLYLCWGYQMLRGALKKEAVSAPQV----- 106

QY 110 TVPDDTPLGGSATDTRNRVRVSVSKORVWVKPMLMAIVLTWLPNAYLDAPVFIGG 169

DB 107 -----ELAKSGRS-----FLKGLLTLANPK---AIIYFGS 134

QY 170 -----VGAQYGDTCRWTFACAFASLIWFLYGFAGALSRPLSSPKWRWINVVAVV 224

DB 135 VFSLFVNDNGVTARWGIFALLIIVETLAWFTVVASLFPQMRGORYOLAKWIDGFAGAL 194

QY 225 MTALAIKML 234

DB 195 FAGFGIHLII 204

RESULT 13

AA01787

ID AAB01787 standard; Protein; 212 AA.

XX AC AAB01787;

XX DT 03-JAN-2001 (first entry)

XX DE Escherichia coli Yeas amino acid excretion protein.

XX KW E. coli; yeas gene; amino acid production; excretion protein gene;

XX KW amino acid excretion protein.

XX OS Escherichia coli.

XX PN EP1016710-A2.

XX PD 05-JUL-2000.

XX PF 17-DEC-1999; 99EP-0125263.

XX PR 30-DEC-1998; 98RU-0124016.

XX PR 09-MAR-1999; 99RU-0104431.

XX XX (AJIN) AJINOMOTO CO INC.

XX XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;

XX PI Tokhmakova IL;

XX DR WPI; 2000-414802/36.

XX DR N4PSDB; AAA52689.

XX PT Increased production of L-amino acids by an Escherichia bacterium

XX PT comprises increasing the expression amount of an L-amino acid excretion

PT protein -

XX Claim 1; Page 21; 29pp; English.

XX The present sequence is the Yeas amino acid excretion protein from
 CC Escherichia coli. This protein is involved in the production of amino
 CC acids, and an increase in its expression leads to an increased
 CC accumulation of amino acids in the cell. In this case, an increase in
 CC lysine, alanine, valine, histidine, isoleucine, glutamic acid and proline
 CC is achieved if multiple copies of its gene are transfected into a
 CC bacterium. The bacterium used is E. coli.

XX SQ Sequence 212 AA;

Query Match 8.5%; Score 101; DB 21; Length 212;

Best Local Similarity 22.3%; Pred. No. 0.0036;

Matches 52; Conservative 28; Mismatches 69; Indels 84; Gaps 9;

QY 12 LLGASLLLSI-GPQNVLVTKQIK---REGLTAVLLVCLISDVFLFIAGTIGVLLISNAA 67

DB 14 LVGAIFIVLVPGEITFLVKNSVSSGMKGGLAACGVFIGDAVLMFLAWA-GVATLIKT 72

QY 68 PIVLDIRMGGIAYLLW-----FAVMAAKDAMTKVAPQIIEETEPVDDTPLGSSA 121

DB 73 PILFNIVRYLGFAPYLLYLGSKILYATLKGN-----EAKSDEPQYGAI 116

QY 122 VATDTRNRVRVSVSKORVWVKPMLMAIVLTWLPNNA---YLDAPVFIGGVAQYGD 178

DB 117 -----FKRALILSLTNPKAILFYVSPFVQFIDVNAHTGIS 152

QY 179 RMIFRA-----GAPAA-----SLIWFPLVFGCA 201

DB 153 PFILAALELVFCVLSFLIISGAFVTVIRTKKLVKGNLSLIGLMFVGFPA 205

RESULT 14

AAU58229

ID AAU58229 standard; Protein; 390 AA.

XX AC AAU58229;

XX DT 27-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #19125.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-199047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX XX (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59589.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for

XX PT vaccinating against and diagnosing infections, especially useful for

XX PT treating acne vulgaris -

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:23:28 ; Search time 17 Seconds
(without alignments)

1112.390 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVIMEIFITGLLGASLLLS.....INVVAVVMTALAIKLMIMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1191	100.0	236	9	US-09-738-626-6955
2	1178	98.9	233	9	US-09-746-660A-52
3	132.5	11.1	226	9	US-09-738-626-3665
4	93.5	7.9	452	10	US-09-815-242-10183
5	93	7.8	459	9	US-09-738-626-6988
6	91.5	7.7	446	9	US-09-738-626-6356
7	91.5	7.7	446	10	US-09-951-780-2
8	91	7.6	624	9	US-09-738-626-4289
9	89.5	7.5	205	10	US-09-927-395-2
10	89.5	7.5	205	10	US-09-847-392-2
11	89.5	7.5	340	10	US-09-815-242-11980
12	86	7.2	295	10	US-09-815-242-10104
13	86	7.2	400	9	US-09-738-626-4488
14	84.5	7.1	415	9	US-09-976-059-20
15	83	7.0	295	10	US-09-815-242-14000
16	82	6.9	708	9	US-09-738-626-5024
17	81.5	6.8	478	9	US-09-738-626-3783
18	81.5	6.8	516	9	US-09-738-626-5086
19	80.5	6.8	483	9	US-09-738-626-3588

20	78.5	6.6	452	10	US-09-815-242-13983	Sequence 13983, A
21	78	6.5	446	9	US-10-217-096-4	Sequence 4, Appli
22	76.5	6.4	632	10	US-09-919-497-75	Sequence 75, Appl
23	76	6.4	323	9	US-09-738-626-5239	Sequence 5239, Ap
24	75.5	6.3	386	10	US-09-925-301-1277	Sequence 1277, Ap
25	75.5	6.3	332	9	US-09-738-626-4243	Sequence 4243, Ap
26	75.5	6.3	501	10	US-09-934-868-56	Sequence 56, Appl
27	74.5	6.3	361	9	US-10-028-072-252	Sequence 252, App
28	74.5	6.3	361	9	US-10-121-049-252	Sequence 252, App
29	74.5	6.3	361	9	US-10-123-904-252	Sequence 252, App
30	74.5	6.3	361	9	US-10-140-470-252	Sequence 252, App
31	74.5	6.3	361	9	US-10-175-746-252	Sequence 252, App
32	74.5	6.3	361	9	US-10-176-918-252	Sequence 252, App
33	74.5	6.3	361	9	US-10-176-921-252	Sequence 252, App
34	74.5	6.3	361	9	US-10-137-865-252	Sequence 252, App
35	74.5	6.3	361	9	US-10-140-474-252	Sequence 252, App
36	74.5	6.3	361	9	US-10-142-431-252	Sequence 252, App
37	74.5	6.3	361	9	US-10-143-114-252	Sequence 252, App
38	74.5	6.3	361	9	US-10-140-002-252	Sequence 252, App
39	74.5	6.3	361	9	US-10-142-419-252	Sequence 252, App
40	74.5	6.3	361	9	US-10-123-262-252	Sequence 252, App
41	74.5	6.3	361	9	US-10-142-423-252	Sequence 252, App
42	74.5	6.3	361	9	US-10-121-050-252	Sequence 252, App
43	74.5	6.3	361	9	US-10-141-755-252	Sequence 252, App
44	74.5	6.3	361	9	US-10-143-032-252	Sequence 252, App
45	74.5	6.3	361	9	US-10-123-108-252	Sequence 252, App

ALIGNMENTS

RESULT 1

```
US-09-738-626-6955
; Sequence 6955, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6955
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6955
```

Query Match 100.0%; Score 1191; DB 9; Length 236;
Best Local Similarity 100.0%; Pred. No. 3.4e-112;
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MVIMEIFITGLLGASLLLSIGPQNVIVIKQIKREGIALLVCLISDFVFIAGTILGV 60
Db 1 MVIMEIFITGLLGASLLLSIGPQNVIVIKQIKREGIALLVCLISDFVFIAGTILGV 60
QY 61 DLLSNAAPIVLDINWGGIAYLLWFAVMAAKDMTKVEAQIIEETPTVDDPTLGG 120
```

Db 61 DLSNAAPVLDIMRWGGIAYLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGS 120
 QY 121 AVATDTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRW 180
 Db 121 AVATDTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRW 180
 QY 181 IFAGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 236
 Db 181 IFAGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 236

RESULT 2

US-09-746-660A-52
 ; Sequence 52, Application US/09746660A
 ; Publication No. US20030049804A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pompejus, Markus
 ; APPLICANT: Kroger, Burkhard
 ; APPLICANT: Schroder, Hartwig
 ; APPLICANT: Zelder, Oskar
 ; APPLICANT: Haberhauer, Gregor
 ; APPLICANT: Kim, Jun-won
 ; APPLICANT: Lee, Heung-Schick
 ; APPLICANT: Hwang, Byung-Joon
 ; TITLE OF INVENTION: CORNEBACTERIUM GLUTAMICUM GENES ENCODING
 ; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
 ; FILE REFERENCE: BGI-121CP2
 ; CURRENT APPLICATION NUMBER: US/09/746,660A
 ; CURRENT FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 09/606740
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: 09/603124
 ; PRIOR FILING DATE: 2000-06-23
 ; PRIOR APPLICATION NUMBER: 60/141031
 ; PRIOR FILING DATE: 1999-06-25
 ; PRIOR APPLICATION NUMBER: 60/142101
 ; PRIOR FILING DATE: 1999-07-02
 ; PRIOR APPLICATION NUMBER: 60/148613
 ; PRIOR FILING DATE: 1999-08-12
 ; PRIOR APPLICATION NUMBER: 60/187970
 ; PRIOR FILING DATE: 2000-03-09
 ; PRIOR APPLICATION NUMBER: DE 19931420.9
 ; PRIOR FILING DATE: 1999-07-08
 ; NUMBER OF SEQ ID NOS: 125
 ; SOFTWARE: Patent in Vers. 2.0
 ; SEQ ID NO 52
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-746-660A-52

Query Match 98.9%; Score 1178; DB 9; Length 233;
 Best Local Similarity 100.0%; Pred. No. 6.7e-111;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIANLVCLISDVFLFIAGTLGVDLL 63
 Db 1 MEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIANLVCLISDVFLFIAGTLGVDLL 60
 QY 64 SNAAPVLDIMRWGGIAYLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGSVA 123
 Db 61 SNAAPVLDIMRWGGIAYLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGSVA 120
 QY 124 TDTNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRWIFA 183
 Db 121 TDTNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRWIFA 180
 QY 184 AGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 236
 Db 181 AGAFAASLIWFLVGFCAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 233

RESULT 3

US-09-738-626-3665
 ; Sequence 3665, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIKO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, NAKO
 ; APPLICANT: SENOH, AKIHIRO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738,626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: Patent in ver. 3.0
 ; SEQ ID NO 3665
 ; LENGTH: 226
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-3665

Query Match 11.1%; Score 132.5; DB 9; Length 226;
 Best Local Similarity 24.2%; Pred. No. 1.2e-05;
 Matches 58; Conservative 45; Mismatches 99; Indels 39; Gaps 12;

QY 7 FITGILLGASLLLSIGPQNVLVIKQIK--REGLIANLVCLISDVFLFIAGTLGVDLLS 64
 Db 6 FLALFLVMTAATASPGDLPQIIRLSAKNRDGVLTAVGIMVGNISWI-IASLLGLSALI 64
 QY 65 NAAPIVLDIMRWGGIAYLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGSVA 124
 Db 65 STYPALNLLQLVGGYLTWVGIGAVRSWTKR-----STQQAADSDQAVENTLVTA 116
 QY 125 DTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQY--GDTG-RW 180
 Db 117 TAAS-----VGWPAIRSGIATNLSNPKAVL----FFGSVPAQFVRPDMWIGM 160
 QY 181 IFAGAFA--AASLIWFLVGFCAALSRPLSS--PKVWRWVNVVAVMTALAKMLMG 236
 Db 161 SIFIGVFLTLTGLLWF--VGF--AVLVKLAAGLTRNGAIIIDLTGTVIFIGLGMFIFEG 216

RESULT 4

US-09-815-242-10183
 ; Sequence 10183, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; TITLE OF INVENTION: Prokaryotes
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078


```
Db 286 SSSSALPNLMRKMEHGVAKPTGIVVPTGYSFNLGDT-----324
QY 148 MAIVLTWLNPNAYLDAPFVIG-----GVGAQYGDTRWIFAA-GAFAASLIWFFPLVGF 199
Db 325 -AIVLTW-----ASIFADANMPMSLGEQVGLLVFMIIASKGAAGVS-----GA 368
QY 200 GAALSRPLSS--PKWVRWINVVAV 223
Db 369 GIATLAAGLSSHRPELLHGVDVIVGI 394
```

RESULT 7

US-09-951-780-2

; Sequence 2, Application US/09951780

; Patent No. US20020106759A1

; GENERAL INFORMATION:

; APPLICANT: FARWICK, MIKE

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE dcta GENE

; FILE REFERENCE: 032301 WD 199

; CURRENT APPLICATION NUMBER: US/09/951,780

; CURRENT FILING DATE: 2001-09-14

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 446

; TYPE: PR1

; ORGANISM: Corynebacterium glutamicum

US-09-951-780-2

```
Query Match 7.7%; Score 91.5; DB 10; Length 446;
Best Local Similarity 24.8%; Pred. No. 0.4;
Matches 51; Conservative 29; Mismatches 71; Indels 55; Gaps 8;
```

```
QY 28 VIKQIKREGIIAVLLVCLISDVFLFIAGTIGVDLLSNAAPILVIDIMRWGGIAYLLWFAV 87
Db 234 VVQLGIL---ILAFYVTCVI-----FIFGVLGAVLVFTGWNIFKLVKYLAKFLLIFAT 285
QY 88 MAKDAMTNKVEAPQIIEETPTVDDTPLGGSAVATDTRNRVRVEVSDKQVRVVKPML 147
Db 286 SSSSALPNLMRKMEHGVAKPTGIVVPTGYSFNLGDT-----324
QY 148 MAIVLTWLNPNAYLDAPFVIG-----GVGAQYGDTRWIFAA-GAFAASLIWFFPLVGF 199
Db 325 -AIVLTW-----ASIFADANMPMSLGEQVGLLVFMIIASKGAAGVS-----GA 368
QY 200 GAALSRPLSS--PKWVRWINVVAV 223
Db 369 GIATLAAGLSSHRPELLHGVDVIVGI 394
```

RESULT 8

US-09-738-626-4289

; Sequence 4289, Application US/09738626

; Publication No. US20020197605A1

; GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI

; APPLICANT: MIZOGUCHI, HIROSHI

; APPLICANT: ANDO, SEIKO

; APPLICANT: HAYASHI, MIKIRO

; APPLICANT: OCHIAI, KEIKO

; APPLICANT: YOKOI, HARUHIKO

; APPLICANT: TATEISHI, NAKO

; APPLICANT: SENOH, AKIHIRO

; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-125

; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18

; PRIOR APPLICATION NUMBER: JP 99/377484

; PRIOR FILING DATE: 1999-12-16

; PRIOR APPLICATION NUMBER: JP 00/159162

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4289
; LENGTH: 624
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4289
```

```
Query Match 7.6%; Score 91; DB 9; Length 624;
Best Local Similarity 20.7%; Pred. No. 0.7;
Matches 54; Conservative 40; Mismatches 71; Indels 96; Gaps 14;

QY 12 LLGASLL-----LSI--GPQNVLVIKQIKREGIIAVLLVCLISDVFLFIAGTIGV 60
Db 112 MLGAALFYGDVVITPAISVLSATEGLTVISPSFER-FILPVSIAVLLA---IFAIQPLGT 167
QY 61 DLLSNA-APILVIDIMRWGGIAYLLWFAVNAAKDAMTNKVEAPQIIEETPTVDDTPLGG 119
Db 168 EKVGKAFGPIM-----LLNEVTLAGL-GIPQIICHPEILQSLSPH-----206
QY 120 SAVATDTRNRVRVEVSDKQVRVVKPMLMAIVLTWLNPNAYLDAPFVIG-----GVGA 172
Db 207 -----WALRLIVA-----EP---FOAFVLLGAVVLTVTGAEA 235
QY 173 QYGDTRWIFAAFAASLIWFFLV-----GFGAALSRPLS-----SPKVMR 215
Db 236 LVADMGHF---GARPIRVAMFCVWMPALLITVLGGALVINQPEAVRPNMFYLAPEGLR 291
QY 216 WINVVAVVMTALAIKMLMG 236
Db 292 IPLVILATITATVIASQAVISG 312
```

RESULT 9

US-09-927-395-2

; Sequence 2, Application US/09927395

; Patent No. US20020058314A1

; GENERAL INFORMATION:

; APPLICANT: LIVSHITS, VITALY ARKADIEVICH

; APPLICANT: ZAKATAEVA, NATALIYA FAVLOVNA

; APPLICANT: ALCOSHIN, VLADIMIR VENIAMOVICH

; APPLICANT: BELAREOVA, ALL VALENTINOVNA

; APPLICANT: TOKMAKOVA, IRINA LVONNA

; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM

; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD

; FILE REFERENCE: 0010-1039-0

; CURRENT APPLICATION NUMBER: US/09/927,395

; CURRENT FILING DATE: 2001-08-13

; PRIOR APPLICATION NUMBER: 09/396,357

; PRIOR FILING DATE: 1999-09-15

; PRIOR APPLICATION NUMBER: RU98118425

; PRIOR FILING DATE: 1998-10-13

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 205

; TYPE: PR1

; ORGANISM: Escherichia coli

US-09-927-395-2

```
Query Match 7.5%; Score 89.5; DB 10; Length 205;
Best Local Similarity 20.6%; Pred. No. 0.23;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY 4 MEIFITGLLLGASLLSIGPQNVLVIKQIKREGIIAVLLVCLISDVFL--LFIAGTIGVD 61
Db 3 LEWFPAYLLTSIILTLSPGSGAINTMTTSL-NHGYPAGGVYCWASDRTGDSYCAGWRGVG 61
QY 62 LLSNAPILVIDIMRWGGIAYLLWFAVNAAKDA-----MTN 96
```

[illegible]

RESULT 10
 US-09-847-392-2
 ; Sequence 2, Application US/09847392
 ; Patent No. US20020102670A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
 ; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
 ; APPLICANT: ALCOSHIN, VLADIMIR VENYAMITOVICH
 ; APPLICANT: BELAREOVA, ALL VALENTINOVNA
 ; APPLICANT: TOKHAKOVA, IRINA LVONNA
 ; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH
 ; TITLE OF INVENTION: ESCHERICHTA COLI RESISTANCE
 ; TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS
 ; FILE REFERENCE: 0010-1039-0
 ; CURRENT APPLICATION NUMBER: US/09/847,392
 ; CURRENT FILING DATE: 2001-05-03
 ; PRIOR APPLICATION NUMBER: 09/396,357
 ; PRIOR FILING DATE: 1999-09-15
 ; PRIOR APPLICATION NUMBER: RU98118425
 ; PRIOR FILING DATE: 1998-10-13
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 2
 ; LENGTH: 205
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli
 US-09-847-392-2

```

Query Match      7.5%; Score 89.5; DB 10; Length 205;
Best Local Similarity 20.6%; Pred. No. 0.23;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY      4 MEIFITGLLLGASLLISIGPONVLVIKQGIKREGILAVLLVCLISDVF--LFIAGTIGVD 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3 LEWFWAYLLSIITLSPGSAINTMTTSL--NHGYPAGGVYCWASDRDGTGSDYCWGRGVG 61
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      62 LLSNAAPVILDIIMRWGGIYALLWFAVMAAKDA-----MTN 96
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 TLFPSRVIAPEVLKWAGAAVYLLWLGICQWRAAGALDLKSLASTOSRRHLFORAFVNLTN 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      97 KVEA-----PQIIEETEPTVPDPTLGGSAVATDTRNRVEVSVQKRWVK 144
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122 PKGISVFLAALFPQIMPQQPOLMOYIVLGVTTIIVVDIIVMTIGYATLAQRIALWIK 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-09-815-242-11980
; Sequence 11980, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11980
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11980

Query Match          7.5%; Score 89.5; DB 10; Length 340;
Best Local Similarity 24.4%; Pred. No. 0.44;
Matches 61; Conservative 29; Mismatches 89; Indels 71; Gaps

Qy 1 MVINEIFITGLLGASLLSIGPQNVLVIKQIKRE-GLIAVLLVCLISDVFLFIAGTLG 59
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 VALPDLASMGQYLGCA-FLGAGLAGIAVFLGQARETGTPVFLVLGAGLSVWLASLTG 171
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 60 VDLLSNAPIVDIMR-W-----GGIAYLLM-----FAVMAAKDAMTNKVEAPOIIE 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 IIVL-NAPPEVFDFRFRHWAAGSLSGSGFALLGWPLGAIGLAGLAAAFALARNALALGOE 230
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 107 TEPTVPDDTLPGGSAAVATDFRNRVREVSVDKQVRVWKPMLMAIVLTWLNPNAYLDAFVF 166
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 I-----GQALGVDIR-----ITWL--LACLAVMLL 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 167 IGVGAQYDGTGRWI-FAAGAFASLIWFPLVGFGAALSRPLSSPKVRWNVVAVVMT 226
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 AG-----AATALAGPIAF---VGLVAPHLARLLIAGPD-QRWILPFSALIAA 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 227 ALAIKMLMG 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 296 GLLIGADILG 305
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-09-815-242-10104
; Sequence 10104, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

matches	34;	conservative	27;	mismatches	13;	indels	98;	gaps	gaps
---------	-----	--------------	-----	------------	-----	--------	-----	------	------

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:21:48 ; Search time 15 Seconds
(without alignments)
462.921 Million cell updates/sec

Title: US-09-105-117k-2
Perfect score: 1191
Sequence: 1 MVIMRIFITGLLGSALLS.....INNVAVVMTALAIAKIMLMG 236

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2.6/prodata/1/iaa/5A COMB.pep:*
2: /cgn2.6/prodata/1/iaa/5B COMB.pep:*
3: /cgn2.6/prodata/1/iaa/6A COMB.pep:*
4: /cgn2.6/prodata/1/iaa/6B COMB.pep:*
5: /cgn2.6/prodata/1/iaa/PCUS COMB.pep:*
6: /cgn2.6/prodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	119.5	10.0	173	4	US-09-134-001C-4026
2	89.5	7.5	205	4	US-09-396-357-2
3	85.5	7.2	224	4	US-09-134-001C-4608
4	82.5	6.9	3491	2	US-07-642-734C-2
5	82.5	6.9	3491	3	US-08-439-009A-2
6	80	6.7	377	4	US-09-105-390-64
7	79	6.6	478	4	US-09-134-001C-4637
8	79	6.6	530	4	US-08-793-044-3
9	79	6.6	4545	2	US-08-804-227C-14
10	79	6.6	4550	2	US-08-804-227C-8
11	79	6.6	4550	2	US-08-804-198-2
12	78.5	6.6	831	2	US-08-677-734A-11
13	78.5	6.6	831	4	US-09-097-053-11
14	76.5	6.4	834	2	US-08-677-734A-9
15	76.5	6.4	834	2	US-08-677-734A-10
16	76.5	6.4	834	4	US-09-097-053-9
17	76.5	6.4	834	4	US-09-097-053-10
18	74.5	6.3	5215	4	US-09-105-537-2
19	73.5	6.2	484	4	US-09-134-001C-5063
20	73	6.1	303	2	US-08-846-762-20
21	73	6.1	303	2	US-08-846-762-91
22	73	6.1	502	4	US-09-134-001C-5674
23	72.5	6.1	488	2	US-08-928-692-10
24	72.5	6.1	488	4	US-09-338-972-10
25	72	6.0	192	1	US-08-112-208C-9
26	72	6.0	192	1	US-08-248-819A-9
27	72	6.0	192	1	US-08-607-269-25

28	72	6.0	192	1	US-08-471-058-13	Sequence 13, Appl
29	72	6.0	192	2	US-08-337-646A-9	Sequence 9, Appl
30	72	6.0	192	2	US-08-856-531-9	Sequence 9, Appl
31	72	6.0	192	2	US-08-856-034-9	Sequence 9, Appl
32	72	6.0	192	3	US-08-471-057-13	Sequence 13, Appl
33	72	6.0	192	4	US-09-127-048-7	Sequence 7, Appl
34	72	6.0	192	4	US-08-927-326-9	Sequence 9, Appl
35	72	6.0	192	5	PCT-US95-04600-25	Sequence 25, Appl
36	72	6.0	245	4	US-08-469-260A-40	Sequence 40, Appl
37	72	6.0	450	3	US-09-045-631-2	Sequence 2, Appl
38	72	6.0	450	4	US-09-158-843A-2	Sequence 2, Appl
39	72	6.0	1447	4	US-09-041-886-25	Sequence 25, Appl
40	72	6.0	1447	5	PCT-US94-05277-2	Sequence 2, Appl
41	72	6.0	1528	1	US-08-463-092B-6	Sequence 6, Appl
42	72	6.0	1528	2	US-08-462-109A-6	Sequence 6, Appl
43	72	6.0	1528	2	US-08-460-907B-6	Sequence 6, Appl
44	72	6.0	1528	3	US-08-463-179A-6	Sequence 6, Appl
45	72	6.0	1528	3	US-08-461-384B-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-134-001C-4026
; Sequence 4026, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4026.
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4026

Query Match 10.0%; Score 119.5; DB 4; Length 173;
Best Local Similarity 23.0%; Pred. No. 3.5e-06;
Matches 46; Conservative 32; Mismatches 83; Indels 39; Gaps 4;
QY 39 IAVLLVCLISDVFLTAGTGLGVDDLNSRAPIVLDIMRWGGIAVLLWFAVWAAKAMTKV 98
Db- 8 LPVITAGUCDTFLIVAILGVSLIISMTLQLFYIIGFLFLMTM-----AWSLWT 60
QY 99 EAPQIIEETPTVPDDTPLGGSAAVATDTNRVRVEVSDKQVRVWVKPMLMAIVLTWLNPN 158
Db 61 EXPNSIEETP-----MSAKQ-----ILFALSVSLNPH 90
QY 159 AVLDPAVFITGGVGAQYGDTRWIFAGAPAAASLIWPLPVGGAALSRPLSPKVRWTN 218
Db 91 AIMDTVGVTGSASVVDGVKVVFSLATISVSWINFLAILGRITGKIDKSGKVIILN 150
QY 219 VV--VAVVMTALAIAKIMLMG 236
Db 151 KVSSVIVIIIGLIILKNIVG 170

RESULT 2
US-09-396-357-2
; Sequence 2, Application US/09396357
; Patent No. 6303348
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA

APPLICANT: ALCOSHIN, VLADIMIR VENTAMIOVICH
APPLICANT: BELAREOVA, ALL VALENTINOVA
APPLICANT: TOKHAKOVA, IRINA LVOVNA
TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L0-HOMOSERINE AND METHOD
TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS
FILE REFERENCE: 0010-1039-0
CURRENT APPLICATION NUMBER: US/09/396,357
CURRENT FILING DATE: 1999-09-15
EARLIER APPLICATION NUMBER: RU98118425
EARLIER FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 2
LENGTH: 205
TYPE: PRP
ORGANISM: Escherichia coli
US-09-396-357-2

Query Match 7.5%; Score 89.5; DB 4; Length 205;
Best Local Similarity 20.6%; Pred. No. 0.015;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY 4 MEIETITGLLIGASLLSIPQNVLVKQIKREGILIAVLLVCLISDFV--LFIAGTIGVD 61
DB 3 LEWFAVLLTSIILTPGSGAINTWTSI--NHGYPAGVTCWASDRGTGDSYCAWGRGVG 61

QY 62 LLSNAPVILDMRWGGIAYLLWFAVMAAKDA-----MTN 96
DB 62 TLFERSVIAPELVKAWAGAVLWIGIQWRAAGAILDKSLASTGSRRLFORAVFVNLN 121

QY 97 KVEA-----PQIETEPTVPDPTPLGGSVATDTRNRVRVSVDKORVWK 144
DB 122 PKSIVFLAALFPQFIMPQPOLMYVLGVTVIVVDIIVMIGYATLAQRIALMIK 176

RESULT 3
US-09-134-001C-4608
Sequence 4608, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4608
LENGTH: 224
TYPE: PRP
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4608

Query Match 7.2%; Score 85.5; DB 4; Length 224;
Best Local Similarity 19.5%; Pred. No. 0.049;
Matches 48; Conservative 42; Mismatches 103; Indels 53; Gaps 8;

QY 3 IMEIFITGLLIGASLLSIPQNVLVKQI---KREGIAYLLV-----CLISDFVLFIA 55
DB 14 LMDGLITFIITLIIIVPGDFIIVMKNTINSKMGFMMAAFGITTTGHILYSSLAIF-- 71

QY 56 GTLGVLLSNAAPVILDMRWGGIAYLLWFAVMAAKDAMTKVAPQIETEPTVPDDT 115
DB 72 ---GIYILSLHFVFTIKILGACIYILGKISLSAHSVDFSKQAL----- 117

QY 116 PLGGSVATDTRNRVRVSVDKORVWKPMMLMAIVLTWLNPNAYLDAFVFIGVGAQYG 175
DB 118 -----ADVRN-----VS-----YITSPQGLSTSLNPKALL-----FYVSIRPQPL 154

QY 176 DTGR-----WIFAAGAFASLIWFLVPGFAAALSRFLSPKVRWVWVVMVMTAL 228
DB 155 SNGNIHMKSEVALFAPSVVWVICLWELFCVFFIQYIKLFSRPRPKAIFDYIVGVFLIGL 214

QY 229 AIKML 234
DB 215 SINLL 220

RESULT 4
US-07-642-734C-2
Sequence 2, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dancakers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 3491 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-642-734C-2

Query Match 6.9%; Score 82.5; DB 2; Length 3491;
Best Local Similarity 28.0%; Pred. No. 5.8;
Matches 52; Conservative 18; Mismatches 53; Indels 63; Gaps 12;

QY 66 AAPIVLDI---MRWGGIAYLLW---FAVMAAKDAMTKVE-----APQIIEETPT 110
DB 53 SAPVVFVFGQAGWAGMAGELLGSRVFA--AAMDACARAFEPVTDWTLAQVLDSPQS 110

QY 111 VPDDTPLGGSVATDTRNRVRVSVDKORVWKPMMLMAI-----VLTWLNPNAYLDAFV- 165
DB 111 -----RRVEV-----VQPALFAVQTSLAALWRSFGVTPDAVVG 143

QY 166 -FTGGVGAQY--GDTGRWTFAGAFASLIW---FPLVFGF---AALSRPLSSPKVR 215
DB 144 HSTGEIAAHVCGAAG-----AADAARAALWREMIPLVNGMDMAVALSADEIEPRIAR 199

QY 216 WINVV 221
DB 200 WDDVV 205

RESULT 6
US-09-105-390-64
; Sequence 64, Application US/09105390
; Patent No. 6288303
; GENERAL INFORMATION:
; APPLICANT: Rodriguez, Raymond
; TITLE OF INVENTION: Rice Beta-Glucanase Enzymes
; TITLE OF INVENTION: and Genes
; NUMBER OF SEQUENCES: 66

RESULT 7

US-09-134-001C-4637
; Sequence 4637, Application US/09134001C
; Patent No. 6380370

; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO THE GENES FOR EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007 -
CURRENT APPLICATION NUMBER: IIS/08/134 001C

: CURRENT FILING DATE: 1998-08-13
 : CURRENT APPLICATION NUMBER: US/00/

;; PRIOR APPLICATION NUMBER: US 60/064,964

;; PRIOR FILING DATE: 1997-11-08

;; PRIOR APPLICATION NUMBER: US 60/055,779

;
PRIOR FILING DATE: 1997-08-14

RESULT 6
 US-09-105-390-64
 Sequence 64
 Application no/00105300
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4637
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4637

Query Match 6.6%; Score 79; DB 4; Length 478;

Best Local Similarity 19.7%; Pred. No. 0.84; Mismatches 80; Indels 112; Gaps 13;

QY 29 IKQIKREGH---IAVLVCLISDVFLFIAGTL-GVDLL-----SNAA-----PIVL 71
DB 102 ISGFVKTGLGRRIALQFVKLFGKTLGLAYSILVGVDLILAPATPSNTARAGGIMPIIK 161
QY 72 DIMRWGG-----IA 80
DB 162 SLSEFGSSPRDGERKMGAFILFTEQGNLITSAMPLTAMAGNPITAOQLAHTAHVQIT 221
QY 81 YLLWFVMAAKADAMTKVAPQIIEETPTVPDDTP-----LGGSVAVATDTRN 128
DB 222 WNNWF-VAALIPGLISLIVVFFIYKLYPPTVKETPNNAKKWATEQLEEMGHMSIA-----E 276
QY 129 RVRVESVDKQVWVKPMLMAVLVTLNPNAYLDAFV-----FIGGVGAQ---YGDTRW 180
DB 277 KMGVGFIIALALWV-----LGSFINVDATLTAFIALALLLTGLAWSIDLINETGAW 329
QY 181 IFAGAFASLIWFLVFGGAALSRLSPKVRWNVVAVVMTALAKMLM 235
DB 330 -----NTLVWFSVLVMAEQNLKLGFP-----WLSKLIAGLNGFSWPVIV 372

RESULT 8

US-08-793-044-3
; Sequence 3, Application US/08793044

; Patent No. 6235497

; GENERAL INFORMATION:

; APPLICANT: Bejanin, Stephane

; APPLICANT: Berrard, Sylvie

; APPLICANT: Cervini, Riccardo

; APPLICANT: Mallet, Jacques

; TITLE OF INVENTION: NOVEL VESICULAR ACETYLCHOLINE CARRIER

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Rhone-Poulenc Rorer Inc.

; STREET: 500 Arcola Road, Mailstop 3c43

; CITY: Collegeville

; STATE: PA

; COUNTRY: USA

; ZIP: 19426

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/793,044

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/FR95/01073

; FILING DATE: 10-AUG-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 94/10044

; FILING DATE: 16-AUG-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Savitzky Esq., Martin F.

; REGISTRATION NUMBER: 29,699

; REFERENCE/DOCKET NUMBER: ST94066-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (610) 454-3816

; TELEFAX: (610) 454-3808

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-793-044-3

Query Match 6.6%; Score 79; DB 4; Length 530;

Best Local Similarity 23.5%; Pred. No. 0.98; Mismatches 31; Mismatches 74; Indels 64; Gaps 12;

QY 9 TGLLGASLLSIGPQNVLVIKQIKREGLIAVLVLC--LISDVFLFIAGTLGVLLSNA 66
DB 7 TQOARAAATKLS---EAVGAALQEFQRQRRLVIVCVALLDNMLYVW-----I 53
QY 67 APIVLDI---MRWGGIAYLLWPAVMAAKDAMTKVAPQIIEET-EPTVPDDTPIGGSAV 122
DB 54 VPIVPDYIAHMEGGS-----EGPTLVSEVWEPTLPPTLANASAY 93
QY 123 AYDT-----RNRVREVSVDKQVWVKPML--MAIVLTWLNPN--NAVLDAPV-- 166
DB 94 LANTSASPTAAGSARSILRPYPTSEDEVKIGVLFASKAILQLLVNPLSGPFIIDRMSYDV 153
QY 167 --IGGVGAQYGDTRWIPA---AGAFASLIWFLVFGGAA 202
DB 154 PLLIGLVWFASTVWFVFAEDYATLFAAR-----SLQGLGSA 190

RESULT 9

US-08-804-227C-14

; Sequence 14, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

; APPLICANT: DeHoff, Bradley S.

; APPLICANT: Kuhstoss, Stuart A.

; APPLICANT: Rostock, Paul R., Jr.

; APPLICANT: Sutton, Kimberly L.

; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: THOMAS G. PLANT 1501

; CITY: INDIANAPOLIS

; STATE: IN

; COUNTRY: USA

; ZIP: 46285

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: ASCII(DOS) Text only

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/804,227C

; FILING DATE: February 21, 1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Plant, Thomas, G.

; REGISTRATION NUMBER: 35,784

; REFERENCE/DOCKET NUMBER: X-8231

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 317-276-2459

; INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4545 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-804-227C-14

Query Match

Best Local Similarity 6.6%; Score 79; DB 2; Length 4545;

Mismatches 58; Conservative 35; Mismatches 93; Indels 84; Gaps 11;

QY	3	IMEIFITGLLLGASLLLSIGPQNVLITKQIKREGLIAVLIVCLISDVFLFIAGTTLGVDL	62
Db	2163	VLRTFVGALLAGGENQFAVRSGV-----HVRRVVPAPVPPASARTVTTAPATAVGEDA	2217
QY	63	LSNAPAPIVLIDNRW-----GGIAYLLWFAVMAKDAMTNKVE-----AP	101
Db	2218	RNDTSDVVPDDRWSGTVLTIGTGTAL--GAQVARRLARSGAARLLVIGRRGAAGPVG	2275
QY	102	QIIEB-----TEPTV-----DDTTP-----LGGSAVATD-----TRN	128
Db	2276	ELVEELTALGSEVAEACDVADRDLAALLAGLPEERPLVAVLHAAGVLDDGVLDLSLTD	2335
QY	129	RVRVEVSDKQVRWYKPMIAVLITWLPNPNALDAFVFIGVGQAQYGDTRWIFAA-----	184

Db 2336 RVDA-VLRDK-----VTARHDELTAFLPLDAFLVLFSSIVGWGNGQAVAAANAA 2387
QY 185 -----GAPASLIWFFPLVFGAAA 203
Db 2388 LDALAQRARRARGARAASIAWGPWAGWAS 2417

RESULT 12
US-08-677-734A-11
; Sequence 11, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-677-734A-11

Query Match 6.6%; Score 78.5; DB 2; Length 831;
Best Local Similarity 20.4%; Pred. No. 2.1;
Matches 53; Conservative 35; Mismatches 93; Indels 79; Gaps 11;

QY 6 IFITGL-----LLGASLLLSIGPQNVLVIKQIKREGLIAVLLVCLISDVFLFI 54
Db 162 VFLSGLMGELKIGLDFLLFGSLIAADVPVAVLAFEEVH-----VNEVLFI 209
QY 55 AGTLGVLLSNAPIVL-----DIMRGGIAYLLWFAVMAAKDAMTKVAPQIIEETEP 109
Db 210 --VFGESLLNDAVTVLYNVFESFVTLGG-----DAVTGDCVKGIVSFFV- 253
QY 110 TVPDDTFLGGSV-----ATDTRNRVRVEVSVDKQVWVKPMLMAVLTLWLPNA 159
Db 254 -----VSLGGTLGVIFAFLLSLVTRTKVRI---IEPGFVVISVLTSEMLSLSA 305
QY + 160 YLDAFVFIG-----GVGAQYGDTRWIFAAAGAAASLIWFFLVFGAAALSRLSS 210
Db 306 IL-AITFCGICCCQKVKRANISEQSAITVRYTMKLSAGAEITIFMFLGISAV-----D 357

QY 211 PKVWRINVVAVMTALAI 230
Db 358 FVIWTNTAFVLLTLVFISV 377

RESULT 13
US-09-097-053-11
; Sequence 11, Application US/09097053
; Patent No. 6392025
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,053
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/677,734
; FILING DATE: 10-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 831 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-097-053-11

Query Match 6.6%; Score 78.5; DB 4; Length 831;
Best Local Similarity 20.4%; Pred. No. 2.1;
Matches 53; Conservative 35; Mismatches 93; Indels 79; Gaps 11;

QY 6 IFITGL-----LLGASLLLSIGPQNVLVIKQIKREGLIAVLLVCLISDVFLFI 54
Db 162 VFLSGLMGELKIGLDFLLFGSLIAADVPVAVLAFEEVH-----VNEVLFI 209
QY 55 AGTLGVLLSNAPIVL-----DIMRGGIAYLLWFAVMAAKDAMTKVAPQIIEETEP 109
Db 210 --VFGESLLNDAVTVLYNVFESFVTLGG-----DAVTGDCVKGIVSFFV- 253
QY 110 TVPDDTFLGGSV-----ATDTRNRVRVEVSVDKQVWVKPMLMAVLTLWLPNA 159
Db 254 -----VSLGGTLGVIFAFLLSLVTRTKVRI---IEPGFVVISVLTSEMLSLSA 305
QY ~ 160 YLDAFVFIG-----GVGAQYGDTRWIFAAAGAAASLIWFFLVFGAAALSRLSS 210

Db 306 IL-AITFCGICCKYKVKANISEQSAATVRYTMKMLASGAETIIFMFLGISAV-----D 357
Qy 211 PKVWRWINVVAVVMTALAI 230
Db 358 PVIWTWNTAFVLLTLVPISV 377

RESULT 14
US-08-677-734A-9
; Sequence 9, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 834 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-677-734A-9

Query Match 6.4%; Score 76.5; DB 2; Length 834;
Best Local Similarity 18.7%; Pred. No. 3.7;
Matches 50; Conservative 41; Mismatches 82; Indels 95; Gaps 11;
Qy 6 IFITGL-----LLGASILLISGPNVLVIKQIKREGLIAVLVCLISDVFLFI 54
Db 165 VFLSGLMGDLQIGLLDFLLFGSLMAADVPVAVLAFVEVH-----VNEVLFI 212
Qy 55 AGTLGVDDLSSNAAPIVL-----DIMRWGG-----IAYLLWFVAVMAAKDAMTNKVEAP 101
Db 213 --VFGESELLDADVTVLVYNVPFESFVALGDNVTGVDCKVGIVSFFVVS-----258
Qy 102 QIIETEPTVPDDTPLGGSV-----ATDTRNRVVEVSVKQVWVKPMLMAIV 151
Db 259 -----LGGTLGVGVVFAFLSLVTRFTKHVRI-----IBPGFVFIISYLSLT 300
Qy 152 LTNLPNAVLDADFVIG-----GVGAQYGDTRWIFAFAGAFASLIWFLVFGGA 202
Db 301 SEMLSLSAIL-AITFCGICCKYKVKANISEQSAATVRYTMKMLASSAETIIFMFLGISAV 359

Qy 203 ALSRPLSSPKVWRWINVVAVVMTALAI 230
Db 360 -----NPIWTWNTAFVLLTLVPISV 380
RESULT 15
US-08-677-734A-10
; Sequence 10, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 834 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-677-734A-10

Query Match 6.4%; Score 76.5; DB 2; Length 834;
Best Local Similarity 18.7%; Pred. No. 3.7;
Matches 50; Conservative 41; Mismatches 82; Indels 95; Gaps 11;
Qy 6 IFITGL-----LLGASILLISGPNVLVIKQIKREGLIAVLVCLISDVFLFI 54
Db 165 VFLSGLMGDLQIGLLDFLLFGSLMAADVPVAVLAFVEVH-----VNEVLFI 212
Qy 55 AGTLGVDDLSSNAAPIVL-----DIMRWGG-----IAYLLWFVAVMAAKDAMTNKVEAP 101
Db 213 --VFGESELLDADVTVLVYNVPFESFVALGDNVTGVDCKVGIVSFFVVS-----258
Qy 102 QIIETEPTVPDDTPLGGSV-----ATDTRNRVVEVSVKQVWVKPMLMAIV 151
Db 259 -----LGGTLGVGVVFAFLSLVTRFTKHVRI-----IBPGFVFIISYLSLT 300
Qy 152 LTNLPNAVLDADFVIG-----GVGAQYGDTRWIFAFAGAFASLIWFLVFGGA 202
Db 301 SEMLSLSAIL-AITFCGICCKYKVKANISEQSAATVRYTMKMLASSAETIIFMFLGISAV 359
Qy 203 ALSRPLSSPKVWRWINVVAVVMTALAI 230

Db 360 -----NPEIWTWNTAFVLLTLVFISV 380

Search completed: April 24, 2003, 18:24:13
Job time : 21 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 6117.94 Seconds
(without alignments)
11293.030 Million cell updates/sec

Title: US-09-105-117K-3

Perfect score: 2374

Sequence: 1 agatactcttggasgaaa.....gtaacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	2374	100.0	2374	1	CGLYSEG X96471 C. glutamicu
C 2	2374	100.0	2374	6	A93933 Sequence 2
C 3	2374	100.0	333150	1	AP005277 Corynebac
C 4	2374	100.0	349980	6	AX127147 Sequence
C 5	993	41.8	993	6	AX063767 Sequence
C 6	993	41.8	993	6	AX244055 Sequence
C 7	870	36.6	870	6	AX123540 Sequence
C 8	822	34.6	822	6	AX063771 Sequence
C 9	822	34.6	822	6	AX244059 Sequence
C 10	798.8	33.6	1771	1	AB083133 Corynebac
C 11	708	29.8	708	6	AX123539 Sequence
C 12	627	26.4	627	6	AX123538 Sequence
C 13	485.2	20.4	993	6	AX067087 Sequence
C 14	170	7.2	345783	1	AP003001 Mesorhizo
C 15	164.4	6.9	8472	1	AE008076 Agrobacte
C 16	164.4	6.9	13051	1	AE009111 Agrobacte
C 17	159.4	6.7	15586	1	SC7H9
C 18	140.6	5.9	33285	1	SC5F8
C 19	135.6	5.7	306250	1	SME591788
C 20	122.8	5.2	14844	1	MBU34849
C 21	122.8	5.2	15239	6	I86263 Sequence 17
C 22	121.2	5.1	14869	1	AE007056 Mycobacte
C 23	121.2	5.1	38500	1	MTCY39
C 24	116.4	4.9	11120	1	RFA311775
C 25	113	4.8	6583	1	SC5H1
C 26	112.6	4.7	10871	1	AE011965
C 27	112.4	4.7	10336	1	AE005530
C 28	112.2	4.7	12070	1	AB011413 Streptomy
C 29	111.2	4.7	266658	1	AP002563 Escherich
C 30	109.8	4.6	1525	1	PSE1GRA
C 31	108	4.5	11024	1	AE000382
C 32	108	4.5	141744	1	ECU28377
C 33	107	4.5	110000	2	AC020884.0
C 34	105.2	4.4	22929	1	AE008808
C 35	105.2	4.4	256050	1	AL627274 Salmonell
C 36	101	4.3	12000	6	AX065613 Sequence
C 37	100.8	4.2	5541	6	AR012071 Sequence
C 38	100.8	4.2	5541	6	AR025196 Sequence
C 39	100.8	4.2	5541	6	AR038653 Sequence
C 40	100.8	4.2	5541	6	AR068500 Sequence
C 41	100.8	4.2	5541	6	AX343066 Sequence
C 42	100.8	4.2	5541	6	AX347518 Sequence
C 43	100.8	4.2	5541	6	AX347541 Sequence
C 44	100.8	4.2	5541	6	AX347564 Sequence
C 45	93.2	3.9	988	6	AX431662 Sequence

ALIGNMENTS

RESULT 1
CGLYSEG/c
LOCUS CGLYSEG 2374 bp DNA linear BCT 18-MAR-2001
DEFINITION C. glutamicum lysE and lysG genes.
ACCESSION X96471
VERSION X96471.1 GI:1729753
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine
exporter protein; Lysine governor.
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 2374)
AUTHORS Vrljic, M., Sahm, H. and Eggeling, L.

TITLE A new type of transporter with a new type of cellular function:
JOURNAL L-lysine export from Corynebacterium glutamicum
MEDLINE Mol. Microbiol. 22 (5), 815-826 (1996)
PUBMED 97126810
REFERENCE 2 (bases 1 to 2374)
AUTHORS Vrljic, M.M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie
FEATURES 1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
source Location/Qualifiers
 1. .2374
 /organism="Corynebacterium glutamicum"
 /strain="R127"
 /db_xref="taxon:1718"
 complement (82. .954)
 /gene="lysG"
 complement (82. .954)
 /gene="lysG"
 /note="Lysine governor"
 /codon_start=1
 /transl_table=11
 /product="Lysine export regulator protein"
 /protein_id="CAA65323.1"
 /db_xref="GI:1729754"
 /db_xref="SWISS-PROT:P94632"
 /translation="MNPIQLTLLSIIDRGSFEGASLALISPSAVSRVKALEHHVG
 RVLVSTQPAKATEAGEVLQVQARMLVLAQSLRLAEIPIITAINADSLSTW
 PFLVFNVAWSGATLTLEDEAHTLSLRRLGVDLVAITREANPVAGCEVVELGTW
 HLAIAITPSLRDVAWVDKLDMAAPVLRGPKVDLDRDLGVRGPRRSIVPS
 ASGFGAIRRGGLGWLLPETQAPMLKAGEVILLDEIPIDTPMYQWRWLRSLARL
 TRAVDAALEGRLP"
 1025. .1726
 /gene="lysE"
 1025. .1726
 /gene="lysE"
 /codon_start=1
 /transl_table=11
 /product="Lysine exporter protein"
 /protein_id="CAA65324.2"
 /db_xref="GI:13397387"
 /db_xref="SWISS-PROT:P94633"
 /translation="MEFITGTLGLGASLLLSIGPONVLVIGIKREGLIIVLLNCLI
 SDVPLFIAGTLGVDLLSNAPVWLDIMRWGGIAYLLWFAVMAKDMTKVAPQIIIE
 ETEPTVDDTPLGGSVAIDITNRVRVSEVDKQVWVRPMLMAIVLITWLNENAYLDA
 FVPIGVGAQYGDGTGRIFAAGAFASLIWFLVLVGFGNAALSRPLSSPKVWRWVNVV
 AVVMTALAIKLMG"
 complement (1723. .2352)
 /gene="orf3"
 complement (1723. .2352)
 /gene="orf3"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA65325.1"
 /db_xref="GI:1729756"
 /db_xref="SPRMBL:P94634"
 /translation="MYALRDIIVASGKALVIGISSYGPETABAAEFMAEEGCLLIHQ
 PSYIINRWVEEFGDNNLQSAANNGLVIAFSPAQLGLLTKDYLIDGIPGSRASQ
 GKSLSRGMVFNNDIMVRKINDIAQERGOSLAQMLAWLREQGEYGADTVTSALIGA
 SSVEQLDNLSDLSLNNLEFSDAELEAIDEISHDAGINIWAKATDSKTREN"
 BASE COUNT 526 a 640 c 648 g 560 t
 ORIGIN

Query Match 100.0%; Score 2374; DB 1; Length 2374;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATACCTCTTGAAGAACCATTGACGATTCGCGTACATGTTTCGCTGCGTAAGGC 60
 DB 2374 AGATACCTCTTGAAGAACCATTGACGATTCGCGTACATGTTTCGCTGCGTAAGGC 2315
 QY 61 TCTTTACGTGGGTATTTCTTCTCCTACGCTCCAGAGCTCACAGCGGAGCGGCTGAGTTCA 120
 |||||||

DB 2314 TCTTTACGTGGGTATTTCTTCTCCTACGCTCCAGAGCTCACAGCGGAGCGGCTGAGTTCA 2255
 QY 121 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCCATCATTAATTAATGTTG 180
 |||||||
 DB 2254 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCCATCATTAATTAATGTTG 2195
 |||||||
 QY 181 GGTGGAGGAACCGGCGGATGACGCTGAGAACTTGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 240
 |||||||
 DB 2194 GGTGGAGGAACCGGCGGATGACGCTGAGAACTTGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 2135
 |||||||
 QY 241 TGGCGTCAATGCTTCTTCTACACACTTTCGCGAGGCGCTGTCTACGGAACAATAATCTCGATGG 300
 |||||||
 DB 2134 TGGCGTCAATGCTTCTTCTACACACTTTCGCGAGGCGCTGTCTACGGAACAATAATCTCGATGG 2075
 |||||||
 QY 301 AATTCAGAGGTTCCCGCGCAGCAGGTAAGTTCCTGTCTGAGGGCATGTTGAACGT 360
 |||||||
 DB 2074 AATTCAGAGGTTCCCGCGCAGCAGGTAAGTTCCTGTCTGAGGGCATGTTGAACGT 2015
 |||||||
 QY 361 GAACAATATTGATGTGTCGCAAGCTCAATGACATCGCCAGGAACGCGGCGCATCT 420
 |||||||
 DB 2014 GAACAATATTGATGTGTCGCAAGCTCAATGACATCGCCAGGAACGCGGCGCATCT 1955
 |||||||
 QY 421 TGGCAGATGGCGCTTTCGATGCTGCTGCGCAGCAAGGAGAGTACGCGCGCGATACCGT 480
 |||||||
 DB 1954 TGGCAGATGGCGCTTTCGATGCTGCTGCGCAGCAAGGAGAGTACGCGCGCGATACCGT 1895
 |||||||
 QY 481 GACCAAGTCAATGATGTTGCTTCGTTCAGTTCAGCAGCTGGGCAACAGCCTTGAATTCAT 540
 |||||||
 DB 1894 GACCAAGTCAATGATGTTGCTTCGTTCAGTTCAGCAGCTGGGCAACAGCCTTGAATTCAT 1835
 |||||||
 QY 541 CAACAACCTTGGAGTTTTCGACCGCGAGTTGGAGGGCATGATGAGATTTCCACGACGC 600
 |||||||
 DB 1834 CAACAACCTTGGAGTTTTCGACCGCGAGTTGGAGGGCATGATGAGATTTCCACGACGC 1775
 |||||||
 QY 601 CGGCATCAACATTTGGCGGAAGCCACCGATTCCAAAACCCCGGAAAACCTAACCCATCAA 660
 |||||||
 DB 1774 CGGCATCAACATTTGGCGGAAGCCACCGATTCCAAAACCCCGGAAAACCTAACCCATCAA 1715
 |||||||
 QY 661 CATCAGTTTGTGTCGCAATGCGGTATCACTGCTGCGCAGCAGCGATTCGATTCAGCGCCA 720
 |||||||
 DB 1714 CATCAGTTTGTGTCGCAATGCGGTATCACTGCTGCGCAGCAGCGATTCGATTCAGCGCCA 1655
 |||||||
 QY 721 CACCTTGGGGCTGGACAGCGGGCGTCACTGCTGCTGCGCGGAAACCCACGCGGGGAA 780
 |||||||
 DB 1654 CACCTTGGGGCTGGACAGCGGGCGTCACTGCTGCTGCGCGGAAACCCACGCGGGGAA 1595
 |||||||
 QY 781 CCAGATCAGGCTTTCGCGCAACCGCGCAGCGGCAAAAATCCACCGTCCCGTGTGCGCGTA 840
 |||||||
 DB 1594 CCAGATCAGGCTTTCGCGCAACCGCGCAGCGGCAAAAATCCACCGTCCCGTGTGCGCGTA 1535
 |||||||
 QY 841 TTGCGCGCGACGCGCGCGCATAAACAAACCGGTCCAAATACGCAATTCGGGTTCACCCA 900
 |||||||
 DB 1534 TTGCGCGCGACGCGCGCGCATAAACAAACCGGTCCAAATACGCAATTCGGGTTCACCCA 1475
 |||||||
 QY 901 GGTACAGCATGTCGCATCAATGGGCTTACCACCAACCGCTGCTTATCGACGCTCAC 960
 |||||||
 DB 1474 GGTACAGCATGTCGCATCAATGGGCTTACCACCAACCGCTGCTTATCGACGCTCAC 1415
 |||||||
 QY 961 CTCACCCCGCACCCGGTTCGGGTGTGTCAGTGGCCACCGCGAAACCGCCCAAGGCGGTGC 1020
 |||||||
 DB 1414 CTCACCCCGCACCCGGTTCGGGTGTGTCAGTGGCCACCGCGAAACCGCCCAAGGCGGTGC 1355
 |||||||
 QY 1021 ATCGGCGCGGTGTTCTTCTTCAATGATGTGTGCGCTTCCACTTGTGTTCTCAT 1080
 |||||||
 DB 1354 ATCGGCGCGGTGTTCTTCTTCAATGATGTGTGCGCTTCCACTTGTGTTCTCAT 1295
 |||||||
 QY 1081 GGGCTCTTTGCTGCTCATCGCGCAACCATACAGTTAAGCGATGCCACCCAGCGCAT 1140
 |||||||
 DB 1294 GGGCTCTTTGCTGCTCATCGCGCAACCATACAGTTAAGCGATGCCACCCAGCGCAT 1235
 |||||||
 QY 1141 AATATCGAGCAGCATCGGCGCGCATTTGACAAAAGATCAAGCCCAAGGTGCCGCGAT 1200
 |||||||
 DB 1234 AATATCGAGCAGCATCGGCGCGCATTTGACAAAAGATCAAGCCCAAGGTGCCGCGAT 1175
 |||||||

QY 1201 GAACAAAGACGTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTCCGCGCTT 1260
Db 1174 GAACAAAGACGTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTCCGCGCTT 1115
QY 1261 AATTCTCTGTTTAAATACCAAGTACATTTCTGGGTCCGATGAGACAGTAAAGACATGGGCC 1320
Db 1114 AATTCTCTGTTTAAATACCAAGTACATTTCTGGGTCCGATGAGACAGTAAAGACATGGGCC 1055
QY 1321 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
Db 1054 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 995
QY 1381 AGTAAATGATTTGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATTCACATGGA 1440
Db 994 AGTAAATGATTTGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATTCACATGGA 935
QY 1441 CACTTTCTCTCAATCATTGATGAGGAGCTTCGAAGGCGCTTCTTAGCCCTTTCCAT 1500
Db 934 CACTTTCTCTCAATCATTGATGAGGAGCTTCGAAGGCGCTTCTTAGCCCTTTCCAT 875
QY 1501 TTCCCTCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
Db 874 TTCCCTCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 815
QY 1561 GGTATCGGCACCCAAACCGGCAAGCAACGAGCGGTGAAGTCTCTTGCAAGCAGC 1620
Db 814 GGTATCGGCACCCAAACCGGCAAGCAACGAGCGGTGAAGTCTCTTGCAAGCAGC 755
QY 1621 GCGGAAATGGTGTGTCGCAAGCAGAACTAAAGCGCACTATCTGAGCGCTTGTCTGA 1680
Db 754 GCGGAAATGGTGTGTCGCAAGCAGAACTAAAGCGCACTATCTGAGCGCTTGTCTGA 695
QY 1681 AATCCCTGTAACCATGCCATCAACGAGATTCGCTATCCATGTTTCTCCCGGTT 1740
Db 694 AATCCCTGTAACCATGCCATCAACGAGATTCGCTATCCATGTTTCTCCCGGTT 635
QY 1741 CAACGAGTGTCTTTGGGTGGAGCAACGCTCACGCTCGCTTGAAGATGAAGCGCA 1800
Db 634 CAACGAGTGTCTTTGGGTGGAGCAACGCTCACGCTCGCTTGAAGATGAAGCGCA 575
QY 1801 CACATATCTCTGCGCGGTGGAGATGTTTGAAGCGGTAAACCGTGAAGCTAATCC 1860
Db 574 CACATATCTCTGCGCGGTGGAGATGTTTGAAGCGGTAAACCGTGAAGCTAATCC 515
QY 1861 CGTGGCGGATGTGAAGTGTAGAACCTTGAACCATGCGCACTTGCGCAATTCGACCCC 1920
Db 514 CGTGGCGGATGTGAAGTGTAGAACCTTGAACCATGCGCACTTGCGCAATTCGACCCC 455
QY 1921 CTCAATGCGGATGCCCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 1980
Db 454 CTCAATGCGGATGCCCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 395
QY 1981 ACCTTTGGTCCCAAGATGTCTTCAAGACCGTGTACCTGGAAGCGGCGCTGATGCTCC 2040
Db 394 ACCTTTGGTCCCAAGATGTCTTCAAGACCGTGTACCTGGAAGCGGCGCTGATGCTCC 335
QY 2041 TGTGGGCGCAGCGCGTATTCATTTGTCGCGTGGGAGGTTTGGTGAGGCAATTGC 2100
Db 334 TGTGGGCGCAGCGCGTATTCATTTGTCGCGTGGGAGGTTTGGTGAGGCAATTGC 275
QY 2101 CCAGGCTTTGTTGGGAGCTTCTTCCGAAACCCAGCTGCTCCCATGTCTAAAGCAGG 2160
Db 274 CCAGGCTTTGTTGGGAGCTTCTTCCGAAACCCAGCTGCTCCCATGTCTAAAGCAGG 215
QY 2161 AGAAGTATCTCTCGATGAGATACCCATGTACACCGATGTATTTGGCAACGATGGCG 2220
Db 214 AGAAGTATCTCTCGATGAGATACCCATGTACACCGATGTATTTGGCAACGATGGCG 155
QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATTCAGG 2280
Db 154 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATTCAGG 95

QY 2281 ATTGCGCCTTACTTACTTCTGAAAAAGTTTACGGTCTTCTCTTCTGCGCGCAGGAA 2340
Db 94 ATTGCGCCTTACTTACTTCTGAAAAAGTTTACGGTCTTCTCTTCTTCTGCGCGCAGGAA 35
QY 2341 TTGGGCGCAGGAGAGTAAACCTTTCAGCAAAATGG 2374
Db 34 TTGGGCGCAGGAGAGTAAACCTTTCAGCAAAATGG 1
RESULT 2
A93933/c A93933 Sequence 2 from Patent WO9723597. 2374 bp DNA linear PAT 22-JAN-2000
LOCUS
DEFINITION
ACCESSION A93933
VERSION A93933.1 GI:6742037
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
1. (bases 1 to 2374)
Vrlijc.M. and Eggeling,L.
PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED
ACTIVITY OF EXPORT CARRIERS
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLIJC MARINA (DE)
FEATURES
source
1. .2374
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 2374; DB 6; Length 2374;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGATACTCTTTGGAGAAACCATGTACCATTTGGTGACATTTGCTCTGGAAGGC 60
Db 2374 AGATACTCTTTGGAGAAACCATGTACCATTTGGTGACATTTGCTCTGGAAGGC 2315
QY 61 TCTTTACGTGGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGCGGCTGAGTTTCA 120
Db 2314 TCTTTACGTGGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGCGGCTGAGTTTCA 2255
QY 121 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATCGTTG 180
Db 2254 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATCGTTG 2195
QY 181 GGTGGAGGAACCGGCGGATGAGGTGAGNACTTTGTCAGTCAGCTGCCCAATGGTCT 240
Db 2194 GGTGGAGGAACCGGCGGATGAGGTGAGNACTTTGTCAGTCAGCTGCCCAATGGTCT 2135
QY 241 TGGCGTCATTTGCTTCTACCACTTTGCGCAGGCGCTGCTCACGGAACAATATCTCGATGG 300
Db 2134 TGGCGTCATTTGCTTCTACCACTTTGCGCAGGCGCTGCTCACGGAACAATATCTCGATGG 2075
QY 301 AATTCAGAGGTTTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360
Db 2074 AATTCAGAGGTTTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 2015
QY 361 GAACATATTTGATATGTTCCGCAAGCTCAATGACATCGCCCAAGGAACGCGGCGCATCT 420
Db 2014 GAACATATTTGATATGTTCCGCAAGCTCAATGACATCGCCCAAGGAACGCGGCGCATCT 1955
QY 421 TGGCAGATGGCGCTTGCATGGGTCTGCGGAGCAAGAGAGTAGTCGCGCGGATACCGT 480
Db 1954 TGGCAGATGGCGCTTGCATGGGTCTGCGGAGCAAGAGAGTAGTCGCGCGGATACCGT 1895
QY 481 GACCAAGTCATTTGTTGTTGCTTCGTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 540
Db 1894 GACCAAGTCATTTGTTGTTGCTTCGTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1835

QY 541 CAACAATTGGAGTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACGAGCG 600
DB 1834 CAACAATTGGAGTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACGAGCG 1775
QY 601 CGGCATCAACATTTGGGCGAAGCCACCGAATCCAAACCCCGGAAACCTTAACCCATCAA 660
DB 1774 CGGCATCAACATTTGGGCGAAGCCACCGAATCCAAACCCCGGAAACCTTAACCCATCAA 1715
QY 661 CATCAGTTTGAATGCCAATCGGTCATCAACATCGTCGACGACGATGATCCAGCGCCA 720
DB 1714 CATCAGTTTGAATGCCAATCGGTCATCAACATCGTCGACGACGATGATCCAGCGCCA 1655
QY 721 CACTTGGGCTCGACGAGCGGCGTGACAAATGCTGTGCGCCGAAACCCACGAGCGGAA 780
DB 1654 CACTTGGGCTCGACGAGCGGCGTGACAAATGCTGTGCGCCGAAACCCACGAGCGGAA 1595
QY 781 CCAGATCAGGCTTGGCGGAAACCGCGGCGGAAATCCACGCTCGGTCGCGGTA 840
DB 1594 CCAGATCAGGCTTGGCGGAAACCGCGGCGGAAATCCACGCTCGGTCGCGGTA 1535
QY 841 TTGCGGCGGACCGCGGATAAACAACGCGGTCCAAATACGCAATTCGGGTTCACCA 900
DB 1534 TTGCGGCGGACCGCGGATAAACAACGCGGTCCAAATACGCAATTCGGGTTCACCA 1475
QY 901 GGTACGACGATTCGATCAACATGAGGCTTTACCGAAACCCGCTGCTTATCGAGCTCAC 960
DB 1474 GGTACGACGATTCGATCAACATGAGGCTTTACCGAAACCCGCTGCTTATCGAGCTCAC 1415
QY 961 CTCACCGGACCGGTTGCGGTCGTCAGTGGCGCACCGCGGAAACCCGAAAGCGGTGTC 1020
DB 1414 CTCACCGGACCGGTTGCGGTCGTCAGTGGCGCACCGCGGAAACCCGAAAGCGGTGTC 1355
QY 1021 ATCGGCGACGGTGGTCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTTGTGTCAT 1080
DB 1354 ATCGGCGACGGTGGTCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTTGTGTCAT 1295
QY 1081 GCGCTCTTTCGCTGCCATGACGGAACCAATACAGGTAAAGCATGCGCCAGCGCAT 1140
DB 1294 GCGCTCTTTCGCTGCCATGACGGAACCAATACAGGTAAAGCATGCGCCAGCGCAT 1235
QY 1141 AATATCGACGACGATCGGCGGCTTGGACAAAGATCAACGCCCAAGGTGCGGCGAT 1200
DB 1234 AATATCGACGACGATCGGCGGCTTGGACAAAGATCAACGCCCAAGGTGCGGCGAT 1175
QY 1201 GAACAAAGACGTCAGAAATTAACACACGAGAAGAACCGCAATGAGTCTTTCGCGCTT 1260
DB 1174 GAACAAAGACGTCAGAAATTAACACACGAGAAGAACCGCAATGAGTCTTTCGCGCTT 1115
QY 1261 AATTCCTTGTATTAATCACCAGTACATCTCGGTCGATGGACAGATTAAGATCGGCCC 1320
DB 1114 AATTCCTTGTATTAATCACCAGTACATCTCGGTCGATGGACAGATTAAGATCGGCCC 1055
QY 1321 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTACCTGATGAGATCTTA 1380
DB 1054 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTACCTGATGAGATCTTA 995
QY 1381 AGTAAATGATGGTCTTAAACATGGTTAATATAGTCTTCATGAACCCCAATCAACTGGA 1440
DB 994 AGTAAATGATGGTCTTAAACATGGTTAATATAGTCTTCATGAACCCCAATCAACTGGA 935
QY 1441 CACTTGTCTCAATCATTAATGAGGACGCTTGAAGGCGCTTCTAGCCCTTTCCAT 1500
DB 934 CACTTGTCTCAATCATTAATGAGGACGCTTGAAGGCGCTTCTAGCCCTTTCCAT 875
QY 1501 TTCCCTCGGCGGTGAGTCAGGCGCTTAAAGCTCTCGAGCATCAGTGGGTGAGTGT 1560
DB 874 TTCCCTCGGCGGTGAGTCAGGCGCTTAAAGCTCTCGAGCATCAGTGGGTGAGTGT 815
QY 1561 GGTATCGGCAACCGGCGCAAGCAACCGAGCGGCTGAGTCTTGTGCAAGCAGC 1620
DB 814 GGTATCGGCAACCGGCGCAAGCAACCGAGCGGCTGAGTCTTGTGCAAGCAGC 755

QY 1621 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGACGCTTGTCTGA 1680
DB 754 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGACGCTTGTCTGA 695
QY 1681 AATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATTCACATGTTTCTTCCCGTGT 1740
DB 694 AATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATTCACATGTTTCTTCCCGTGT 635
QY 1741 CAAAGAGTAGTCTTGGGGTGGAGCAACGCTACGCTGCGCTTGGAGATGAAGCGCA 1800
DB 634 CAAAGAGTAGTCTTGGGGTGGAGCAACGCTACGCTGCGCTTGGAGATGAAGCGCA 575
QY 1801 CACATATCTTCTGCGCGGTGGAGATGTTTAAAGCGGTAAACCGTGAAGCTAATCC 1860
DB 574 CACATATCTTCTGCGCGGTGGAGATGTTTAAAGCGGTAAACCGTGAAGCTAATCC 515
QY 1861 CCGTGGCGGATCTGAAGTAGTAGAATCTTGAACCATGCGCCACTTGGCCATTCGAACCC 1920
DB 514 CCGTGGCGGATCTGAAGTAGTAGAATCTTGAACCATGCGCCACTTGGCCATTCGAACCC 455
QY 1921 CTCAATGCGGATGCTACATGTTGATGGAAACTAGATGGGTGCGATGCCGTCTT 1980
DB 454 CTCAATGCGGATGCTACATGTTGATGGAAACTAGATGGGTGCGATGCCGTCTT 395
QY 1981 AGCTTCGCTCCAAAGATGTCCTTCAAGACCGTCACTGACGCGGCGCTCGATGGTCC 2040
DB 394 AGCTTCGCTCCAAAGATGTCCTTCAAGACCGTCACTGACGCGGCGCTCGATGGTCC 335
QY 2041 TGTGGGCGCGACGCGCTATCCATTTGTCGCTGCGGGAAGGTTTGTGAGGCAATTCG 2100
DB 334 TGTGGGCGCGACGCGCTATCCATTTGTCGCTGCGGGAAGGTTTGTGAGGCAATTCG 275
QY 2101 CCGAGGCTTGTGTTGGGACCTTCTCCGAAACCAAGCTGCTCCCATGCTAAAAGCAGG 2160
DB 274 CCGAGGCTTGTGTTGGGACCTTCTCCGAAACCAAGCTGCTCCCATGCTAAAAGCAGG 215
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTCACACCGATGATTTGGCAAGTGGCG 2220
DB 214 AGAAGTGATCTCTCGATGAGATACCCATTCACACCGATGATTTGGCAAGTGGCG 155
QY 2221 CCGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATCGAGG 2280
DB 154 CCGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATCGAGG 95
QY 2281 ATTGGGCTTATGTTACTTCTGAAAAGTTTCAAGGTTTTCATCTTCCGCGCAGGAA 2340
DB 94 ATTGGGCTTATGTTACTTCTGAAAAGTTTCAAGGTTTTCATCTTCCGCGCAGGAA 35
QY 2341 TTGGGCGGCGAGATTAACACCTTCAGCAATGG 2374
DB 34 TTGGGCGGCGAGATTAACACCTTCAGCAATGG 1

RESULT 3

AP005277

LOCUS

DEFINITION

AP005277

Corynebacterium glutamicum ATCC 13032 DNA, linear

BCT 08-AUG-2002

4/10.

ACCESSION

AP005277

BA000036

VERSION

AP005277.1

GI:21323710

KEYWORDS

SOURCE

ORGANISM

Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.

Corynebacterium glutamicum ATCC 13032

Bacteria; Actinobacteria; Actinobacteriales;

Corynebacterineae; Corynebacteriaceae; Corynebacterium.

REFERENCE

1

Nakagawa,S.

Complete genomic sequence of Corynebacterium glutamicum ATCC 13032

Unpublished

JOURNAL

2 (bases 1 to 333150)

REFERENCE

Nakagawa,S.

Direct Submission

Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.

Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida,
Tokyo 194-8533, Japan (E-mail: snakagawa@xanagen.com,
Tel:81-44-829-3031, Fax:81-44-813-1651)
This sequence is conducted by collaboration of Kyowa Hakko Kogyo
Co. Ltd. And Kitagato University.

FEATURES

```

source
  1..333150
    /organism="Corynebacterium glutamicum ATCC 13032"
    /strain="ATCC 13032"
    /db_xref="taxon:196627"
    /note="ATCC 13032"
    /complement(63..1292)
    /gene="Cg10944"
    /complement(63..1292)
    /gene="Cg10944"
    /codon_start=1
    /transl_table=11
    /product="Hypothetical membrane protein"
    /protein_id="BAB98337.1"
    /db_xref="GI:21323711"
    /translation="MTNPTERNARLLIWANGLQNGDOIIVAAKTVLPWLLQAAGAPG
    FLIALLVPIREAGSMPLQAAITGMVLRQTSRKVYVIGSGQFVSALGIGVAAFLRG
    WAGITVIVLLAALSIFRSMCSASKDVQGVKISKRGGLVTRATVIGVWGLVAGL
    AAIKFGSHFTVLAADVIASSFSLFASIVFAIEYAKPATPKNAPSANPWRRCI
    AAKDKAFRRFVLVRSMLVTALSTAFIVALAESGNSIDSLGFFLIASGLASWYGI
    RISGIWSDHSSKNMAGGALFGSIVLILVLSAFAPAQINTLVPLSPFLITLAHTA
    IRVARKTYVMDAEGDQRTRYVADANTLMGVLLIIVGSLGFTAFISGEAALLFLAAI
    GLAGTISARGLKEVSAG"
    1382..2917
    /gene="Cg10945"
    1382..2917
    /gene="Cg10945"
    /codon_start=1
    /transl_table=11
    /product="Putative multicopper oxidases"
    /protein_id="BAB98338.1"
    /db_xref="GI:21323712"
    /translation="MPENLRRTFFKGAVLAATVGAQVLVACSSDDVRYGGEPRPL
    PIPADLGTREGSSVHFALAEQGESQILPDVTTKTWFGNTHLGPTLVVKKGDVHV
    DVINLDEMTVHHMGMLPAIADGGPHSPIGQGTWSPPTVANDAAATLVKHPHTHG
    LTCGLHAYRGLAGMLIVEDEATDKLDLPREYGVDDIPLVMDHRFLEDGSLDEEDLPDL
    GLGDDPTANGIINAHFATRRVRFRVLNGSNRFLNLAFSDDRTTFQVIASDSGLLD
    EPQDRTILAIGPGEWEIWELEPGEDVTLESVGFEDNYGVPDPEFVDFGMSDSFQL
    LTTGPSDDAAQAPALFGLVAKSTEPDVIDATERPTIMNTFSINDIQMDQRVDVVID
    HDQPEYVITNDNSDPHNFHVDARFKVLKEGTDVFLPNDGKDKTGLPGATATL
    AVFEFGHPDPQWPMYHCHMLYHEDQGMGQFVIVPEGDEPAVLGSGTSSIDSAGG
    HAH"
    3063..3983
    /gene="Cg10946"
    3063..3983
    /gene="Cg10946"
    /note="PF00005:ABC transporter
    TIGR00960:3a0501s02: Type II (General) Secretory Pathway
    (LISP) Family protein
    TIGR00968:3a0106s01: sulfate transport system permease
    protein
    TIGR01166:cbiO: cobalt transport protein ATP-binding
    subunit
    TIGR01184:ntrCD: nitrate transport ATP-binding subunits C
    and D
    TIGR01186:proV: glycine betaine/L-proline transport ATP
    binding subunit
    TIGR01187:potA: spermidine/putrescine ABC transporter
    ATP-binding subunit
    TIGR01188:draA: daunorubicin resistance ABC transporter
    ATP-binding subunit
    TIGR01189:ccmA: heme exporter protein CcmA"
    /codon_start=1
    /transl_table=11
    /product="ABC-type transporter, ATPase component"
    /protein_id="BAB98339.1"
    /db_xref="GI:21323713"
    /translation="WTTTPAIDVTDLVRTYGYDTAVKGLNFHFVQRGEVFLGTINGAG
    KTSTLIEIGLSAPSSGTVRISGLDPVADRAILREIPGLMQSGGLPSQLTVAETMDM
    WHGTCTPRAIKVDLADVLLHRENVKVGSGGEQRRLDLACALGDPSTILFLDEPT
    TGLDPSRRHTWQLLDLQKQGVFWMLTHVLEAEFLCDRIATNAGEIAVEGFLDE
    LVARKSIIISFVLRGQVLEFVLSGAEIIRDNHVRVIAITTLQQTLSILTWAAETGI
    ALGGAAPKATPILSVFMDIASLENTSLQTA"
    4003..4836
    /gene="Cg10947"
    4003..4836
    /gene="Cg10947"
    /codon_start=1
    /transl_table=11
    /product="ABC-type transporter, permease components"
    /protein_id="BAB98340.1"
    /db_xref="GI:21323714"
    /translation="MTTSTARGLQHATPERKTSFFKTSLFKAELQFRNKTLFLFM
    ATVPFGVPIPLLFLIGNGAAESANSFYFVNTILLFVQFTVLSMATIRDERVLKR
    LRTGEARDIDIIGAIICFFGALLTLFTVVIIPLLMVLGAPAPINLVPVFAVLIGLLI
    CSALMDSGFTNAEAQMTMPVFMLAGGLGSIIRFVFGDSIVADTLATVPFAAIS
    DLVQIGWAGATPADSVGCGEAFQIDMLIPLIGLAAWTAATAAANAANRYMRWDSY
    R"
    4879..6138
    /gene="Cg10948"
    4879..6138
    /gene="Cg10948"
    /EC_number="2.7.3.-"
    /codon_start=1
    /transl_table=11
    /product="Two-component system, sensory transduction
    histidine kinases"
    /protein_id="BAB98341.1"
    /db_xref="GI:21323715"
    /translation="MNKDFWTAGTARFSGVSLASPVTAPLNSWPERLPNLAKYTL
    YTRVSLQAIPIVLLSAYPLGIVANAGTLNPSFVLLGFSVILLIVTVLYVYQFSLNS
    HPRRSVQFFFTIGLAVLVVVSVIQIPLGNMNDNRATLIFTLCVFLLSLAYI
    PMNRYRWMLIAMSVLAWMTSTTDDYLSALWVVPPLMAGTVRLSVMTVDVMKEVRS
    RELEASERFERFAQELHDTLQHLAAMSVKESALALAKRGDDRLLENRELQK
    LTRTSMSEMRDVRVGYRTNLAITEGAKSLADAHILSVIGTTSVQVPAHRELCAW
    LVRSATNLIHLSDATDNLTLSTSEVRMDNNGVKDIGRLSGLSALRSRAESAGMTL
    IVSDEDQPSVRMLINAPANTPAKEA"
    6139..6750
    /gene="Cg10949"
    6139..6750
    /gene="Cg10949"
    /note="PF00072:Response regulator receiver domain
    PF00196:Bacterial regulatory proteins, luxR family"
    /codon_start=1
    /transl_table=11
    /product="Two-component system, response regulators
    consisting of a CheY-like receiver domain and a HTH
    DNA-binding domain"
    /protein_id="BAB98342.1"
    /db_xref="GI:21323716"
    /translation="MISISADDEALIASLTLISLEPDLIDVPTAGSGBELIETWA
    DPSNRDVCVLDLQDGIDIDTATRLMETTPDLAVLIVTSHPARPLKRALAAGVLG
    FLPTSTADBEATAIRTVHAGRRVIDPELAWTISAGESPLTNEEEVLELAGQGLSA
    BEIAVAHAPGCTTRNYLSQAMTKVQGNRFEATFRARELGWL"
    complement(6781..6987)
    /gene="Cg10950"
    complement(6781..6987)
    /gene="Cg10950"
    /codon_start=1
    /transl_table=11
    /product="Hypothetical protein"
    /protein_id="BAB98343.1"
    /db_xref="GI:21323717"
    /translation="MKHYWSMYFCAGPLNTPHTKXVWFMDIVSVNHWFEWSTNNWVN
    MTGLPRETFVPAFVGSTVLVLSF"
    7048..8205
    /gene="Cg10951"
    7048..8205
    /gene="Cg10951"
    /codon_start=1
    /transl_table=11

```

/product="tRNA-processing ribonuclease BN"
 /protein_id="BAB98344.1"
 /db_xref="GI:21323718"
 /translation="MTAWLCSPSPQVQDEYDMWNGVQVQEHLDATLIAADPHGNPEN
 SGRKERLNFQWKYKALNRTVRDVRPDGLLDLAALITFPFSLISIAPAVLGYSLITIP
 LADSTFELNLNDEVNOYVPEQSHVNGVIDSIAGSAAQGVGVAVGVIPTALWSS
 AYVRAFCRCANAVGRSEGRTLIKRWMLLFNLALLGIIILISVWLNFLVWGFIF
 APIAEPLHLTNLFLTDREPMFIWVRFPIVGVLMFVATLYIYNAPNRPWFRWL
 SLGSLFATVIGLILNGLNFYFTLFAAFSSYGAAGSLAVFIALVFNICLIIGLKID
 VEISRAKIQAQGMPADYSLVPPRSIERKVKQROQLMDQAAIREESN"
 8672..10114
 /gene="Cg10952"
 8672..10114
 /gene="Cg10952"
 /note="PF00005:ABC transporter
 PF00664:ABC transporter transmembrane region."
 /codon_start=1
 /transl_table=11
 /product="ABC-type multidrug/protein/lipid transport
 system, ATPase component"
 /protein_id="BAB98345.1"
 /db_xref="GI:21323719"
 /translation="MRTHTASSILRRIRRRQKGVAFGLFMWOLSEALVPIAI
 GLIVDAVLTKDLRLVGLVAPVLFVLSYREGSRALNRAVNFESHARVEVAD
 HALKNDLPRLVPEVMSRSTADSDSTRIFGQITGVSAATGFLGAATYLLISWLV
 GLLVLVPIISGVVALASKGISKRSVTOQEKLAESGAQSDIMMLRLVRIKAIIGERW
 AVKTFEKSQAARAADVAVASGVAGVIGELSIIVNLAAVLLLAGWRVTTGELPGQ
 LIAIVGVAVLSEPIRLSNSINASAIHGAERVANFLNLDSEQAQVSESSINDGE
 FLIVVPASTPLHGNLILATPHAADIFEGTLRSNLSMHNEDNVPIDPOVIRASGLTDI
 IEVDGDAVPRDGTGNSLGGORORVALRALHADAEVLVLMPTSAVDSVTEVSLAQG
 IKQLRAGKTTIVSSSPAPFNLDLRVISHV"
 10107..11843
 /gene="Cg10953"
 10107..11843
 /gene="Cg10953"
 /note="PF00005:ABC transporter

Query Match 100.0%; Score 2374; DB 1; Length 333150;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATACCTCTTTGAGAAACCATGTACGATTCGCGACATTTGTCGCTCGAAAGGC 60
 DB 327645 AGATACCTCTTTGAGAAACCATGTACGATTCGCGACATTTGTCGCTCGAAAGGC 327704
 QY 61 TCTTTACGTGGTATTTCTTCTACGTCCAGAGCTCACAGCGGAGCGGCTCGAGTTTCAT 120
 DB 327705 TCTTTACGTGGTATTTCTTCTACGTCCAGAGCTCACAGCGGAGCGGCTCGAGTTTCAT 327764
 QY 121 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCAGCTATTCATATTAATCGTTG 180
 DB 327765 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCAGCTATTCATATTAATCGTTG 327824
 QY 181 GGTGAGGACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTCCCAACATGTCT 240
 DB 327825 GGTGAGGACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTCCCAACATGTCT 327884
 QY 241 TGGCGTCAATGCTTTCTCACCACCTTGGCGAGGGCTTGTCTACGGACAAATATCTCGATGG 300
 DB 327885 TGGCGTCAATGCTTTCTCACCACCTTGGCGAGGGCTTGTCTACGGACAAATATCTCGATGG 327944
 QY 301 AATTCCAGAGGGTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 360
 DB 327945 AATTCCAGAGGGTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 328004
 QY 361 GAACAAATATTGATATGTCGCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 420
 DB 328005 GAACAAATATTGATGTCGCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 328064
 QY 421 TGGCGAGATGCGCTTGATGGGTGCTGCGCGAGCAGGAGTACGCGCGGATACCGT 480
 DB 328065 TGGCGAGATGCGCTTGATGGGTGCTGCGCGAGCAGGAGTACGCGCGGATACCGT 328124

QY 481 GACCAGTCATTGATTGTTGGTCTTGTGTCAGTCAGCTGGCAACAGAGCTTGAATTCAT 540
 DB 328125 GACCAGTCATTGATTGTTGGTCTTGTGTCAGTCAGCTGGCAACAGAGCTTGAATTCAT 328184
 QY 541 CAACAACCTTGGAGTTTCTGACCGCGAGTTGGAGGCGATTCGATGAGATTTCCTCCAGAGCG 600
 DB 328185 CAACAACCTTGGAGTTTCTGACCGCGAGTTGGAGGCGATTCGATGAGATTTCCTCCAGAGCG 328244
 QY* 601 CGGCATCAACATTTTGGGCGAAGCCACCGATTCCAAATCCCGCGAAACCTAAACCCATCAA 660
 DB 328245 CGGCATCAACATTTTGGGCGAAGCCACCGATTCCAAATCCCGCGAAACCTAAACCCATCAA 328304
 QY 661 CATCAGTTTGTATGGCCCAATGCGGTTCATCACTGCGCAGCAGCGATTGATCCAGGCGCA 720
 DB 328305 CATCAGTTTGTATGGCCCAATGCGGTTCATCACTGCGCAGCAGCGATTGATCCAGGCGCA 328364
 QY 721 CACCTTGGGGCTGGACAGCGGGCGGTGACAAATGCTGTCGCGCGAAACCCACAGCGGGAA 780
 DB 328365 CACCTTGGGGCTGGACAGCGGGCGGTGACAAATGCTGTCGCGCGAAACCCACAGCGGGAA 328424
 QY 781 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAATCCACCGTTCGGTGTGCGCGTA 840
 DB 328425 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAATCCACCGTTCGGTGTGCGCGTA 328484
 QY 841 TTGGCGCGCGAGCGCGCGCATTAACACAAACCGGTCCAAATACGATTCGGGTTCACCA 900
 DB 328485 TTGGCGCGCGAGCGCGCGCATTAACACAAACCGGTCCAAATACGATTCGGGTTCACCA 328544
 QY 901 GGTTCAGCAGATTGGCCATCAACATGCGGTTCACCAACCCGCTGCTTATCAGCGCTAC 960
 DB 328545 GGTTCAGCAGATTGGCCATCAACATGCGGTTCACCAACCCGCTGCTTATCAGCGCTAC 328604
 QY 961 CTCACCCGCGACCCCGTTCGCGGTGTCAGTGGCCACCGCGCAACCGCCCAAGGCGTGC 1020
 DB 328605 CTCACCCGCGACCCCGTTCGCGGTGTCAGTGGCCACCGCGCAACCGCCCAAGGCGTGC 328664
 QY 1021 ATCGGCGACGGTGGTTCGTTCTTCAATGATCTGTGCGGCTTCACCTTGTGTCAT 1080
 DB 328665 ATCGGCGACGGTGGTTCGTTCTTCAATGATCTGTGCGGCTTCACCTTGTGTCAT 328724
 QY 1081 GGGCTCTTTGCTGCCATGACCGGCAACCATTAACAGGTAAAGCATGCCACCCAGCGCAT 1140
 DB 328725 GGGCTCTTTGCTGCCATGACCGGCAACCATTAACAGGTAAAGCATGCCACCCAGCGCAT 328784
 QY 1141 AATATCGACGACGATCGCGCGCATTTGACAAAGATCAACGCCCAAGTGC CGCGCAT 1200
 DB 328785 AATATCGACGACGATCGCGCGCATTTGACAAAGATCAACGCCCAAGTGC CGCGCAT 328844
 QY 1201 GAACAAAAGAGCTCAGAAATTAACACACAGAGAGAACCGCAATGAGTCTTCGCGCTT 1260
 DB 328845 GAACAAAAGAGCTCAGAAATTAACACACAGAGAGAACCGCAATGAGTCTTCGCGCTT 328904
 QY 1261 AATTCCTTGTTAATCACCAGTACATTCCTGCGGTCCGATGGACAGTAAAGACTTGGCCCC 1320
 DB 328905 AATTCCTTGTTAATCACCAGTACATTCCTGCGGTCCGATGGACAGTAAAGACTTGGCCCC 328964
 QY 1321 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
 DB 328965 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTA 329024
 QY 1381 AGTAAATGATTTGGTTCTTAACATGTTTAAATATAGTTCATGAACCCCATTCACCTGGA 1440
 DB 329025 AGTAAATGATTTGGTTCTTAACATGTTTAAATATAGTTCATGAACCCCATTCACCTGGA 329084
 QY 1441 CACTTTGCTCTCAATCATTTGATGAAGCAGCTTCGAAGCGGCTCTTAGCCCTTTCCAT 1500
 DB 329085 CACTTTGCTCTCAATCATTTGATGAAGCAGCTTCGAAGCGGCTCTTAGCCCTTTCCAT 329144
 QY 1501 TTCCCTCTCGCGGTCAGTCAGCGGTTAAAGCTCTCGAGCATACAGTGGTTCAGTGT 1560
 DB 329145 TTCCCTCTCGCGGTCAGTCAGCGGTTAAAGCTCTCGAGCATACAGTGGTTCAGTGT 329204
 QY 1561 GGTATCGCGCACCCACCGGCCCAAGCAACCGAGCGGTGAAGTCTTGTGCAAGCAGC 1620

Db 329205 GGTATCGCGACCCAAACCGGCCAAAGCAAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 329264
 QY 1621 GCGGAAATGCTGCTGCAAGCAGAACTTAAGCGCACTATCTGAGCCCTTGTGA 1680
 Db 329265 GCGGAAATGCTGCTGCAAGCAGAACTTAAGCGCACTATCTGAGCCCTTGTGA 329324
 QY 1681 AATCCCGTTAAACCATCGCCATCAACGAGATTTGGCTATCACATGGTTTCTCCCGTGT 1740
 Db 329325 AATCCCGTTAAACCATCGCCATCAACGAGATTTGGCTATCACATGGTTTCTCCCGTGT 329384
 QY 1741 CAACGAGTACTTCTTGGGTGAGCAACGCTCACGCTGCTTGGAGATGAAGCGCA 1800
 Db 329385 CAACGAGTACTTCTTGGGTGAGCAACGCTCACGCTGCTTGGAGATGAAGCGCA 329444
 QY 1801 CACATATCTTCTGCGCGGTGAGATGTTTGAAGAGCGGTAAACCGCTGAAGCTAATCC 1860
 Db 329445 CACATATCTTCTGCGCGGTGAGATGTTTGAAGAGCGGTAAACCGCTGAAGCTAATCC 329504
 QY 1861 CGTGGCGGATGTGAAGTAGTAGAATTTGGAACCATCGCCACTTGGCCATTTGCAACCCC 1920
 Db 329505 CGTGGCGGATGTGAAGTAGTAGAATTTGGAACCATCGCCACTTGGCCATTTGCAACCCC 329564
 QY 1921 CTCAATTCGGGATGCTTACATGTTGTGATGGGAACTAGATGGGCTGCGATGCCGCTTT 1980
 Db 329565 CTCAATTCGGGATGCTTACATGTTGTGATGGGAACTAGATGGGCTGCGATGCCGCTTT 329624
 QY 1981 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTGACCTCGAGCGGCGCTGATGGTCC 2040
 Db 329625 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTGACCTCGAGCGGCGCTGATGGTCC 329684
 QY 2041 TGTGGGCGCAGGCGGTATCTCAATTTGCTCGGCGGAAGTGTGCTGAGGCAATTCG 2100
 Db 329685 TGTGGGCGCAGGCGGTATCTCAATTTGCTCGGCGGAAGTGTGCTGAGGCAATTCG 329744
 QY 2101 CCGAGGCTTGGTGGGCACTTCTTCCGAAACCCAGTGTCTCCCATGCTTAAAGCAGG 2160
 Db 329745 CCGAGGCTTGGTGGGCACTTCTTCCGAAACCCAGTGTCTCCCATGCTTAAAGCAGG 329804
 QY 2161 AGAAGTATCTCTCGATGATACCATTTGACACACCATGATTTGGCAAGTGGCG 2220
 Db 329805 AGAAGTATCTCTCGATGATACCATTTGACACACCATGATTTGGCAAGTGGCG 329864
 QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGCGCGTGTGATGACAGCAATTCGAGG 2280
 Db 329865 CTTGGAATCTAGATCTCTAGCTAGACTCACAGCGCGTGTGATGACAGCAATTCGAGG 329924
 QY 2281 ATTGCGGCTTACTTCTGAAAAGTTTCAAGTTTTCATTTCTTCCCGCAGGAA 2340
 Db 329925 ATTGCGGCTTACTTCTGAAAAGTTTCAAGTTTTCATTTCTTCCCGCAGGAA 329984
 QY 2341 TTGGCCAGGCGAGATTAACCTTCAAGCAATGG 2374
 Db 329985 TTGGCCAGGCGAGATTAACCTTCAAGCAATGG 33018

RESULT 4

LOCUS AX127147 349980 bp DNA linear PAT 11-MAY-2001
 DEFINITION Sequence 7063 from Patent EP1108790.
 ACCESSION AX127147 AX114121
 VERSION AX127147.1 GI:14041135

KEYWORDS

ORGANISM

Corynebacterium glutamicum.
 Corynebacterium glutamicum
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 Corynebacterium.

REFERENCE

1 (bases 1 to 349980)
 Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
 Novel polynucleotides
 Patent: EP 1108790-A 7063 20-JUN-2001;

FEATURES
 source

KYOWA HAKKO KOGYO CO., LTD. (JP)
 Location/Qualifiers
 1..349980
 /organism="Corynebacterium glutamicum"
 /db_xref="taxon:1718"
 /note="Seq 1 to long (3,309,400) split in 11, seq 7063
 1.200.001 1.549.980"
 BASE COUNT 80727 a 91049 c 97618 g 80586 t
 ORIGIN

Query Match 100.0%; Score 2374; DB 6; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATACCTCTTTGGAGAAACCATGTACGATTTGGCTGACATTTGGCTCTGGAAAGCG 60

Db 127595 AGATACCTCTTTGGAGAAACCATGTACGATTTGGCTGACATTTGGCTCTGGAAAGCG 127654

QY 61 TCTTTACGTGGGTATTTCTTCTTACGGTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 120

Db 127655 TCTTTACGTGGGTATTTCTTCTTACGGTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 127714

QY 121 GCGGAGGAGGGCTGCGCTTCTGATTCATCAGCAAGCTATTCATTAATTCGTTG 180

Db 127715 GCGGAGGAGGGCTGCGCTTCTGATTCATCAGCAAGCTATTCATTAATTCGTTG 127774

QY 181 GGTGGAGGAACCGGCGATGACGGTGGAGAACTTTGTCAGTCAGCTGCCAACAAATGGTCT 240

Db 127775 GGTGGAGGAACCGGCGATGACGGTGGAGAACTTTGTCAGTCAGCTGCCAACAAATGGTCT 127834

QY 241 TGGCGCTCATTTCTTCTACCACTTTCGCGAGGCTGCTCAGGCAAAATATCTCGATGG 300

Db 127835 TGGCGCTCATTTCTTCTACCACTTTCGCGAGGCTGCTCAGGCAAAATATCTCGATGG 127894

QY 301 AATTCAGAGGGTTCCTCGCGCCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360

Db 127895 AATTCAGAGGGTTCCTCGCGCCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 127954

QY 361 GAACATATGATATGCTCCGCAAGCTCAATGACATCGCCAGGAACGCGGGCAGTCACT 420

Db 127955 GAACATATGATATGCTCCGCAAGCTCAATGACATCGCCAGGAACGCGGGCAGTCACT 128014

QY 421 TGGCAGATGCGCTTGCATGCTGCTGCGCAGCAAGGAGAGTACGGCGCGATACCGT 480

Db 128015 TGGCAGATGCGCTTGCATGCTGCTGCGCAGCAAGGAGAGTACGGCGCGATACCGT 128074

QY 481 GACCACTGATTTGCTGCTTCTGATGAGCAGCTGGAACAACAGCTTGAATTCAT 540

Db 128075 GACCACTGATTTGCTGCTTCTGATGAGCAGCTGGAACAACAGCTTGAATTCAT 128134

QY 541 CAACACTTGGAGTTTCTGACGCGGATTTGGAGGGGATCGATGAGATTTCCACGACGC 600

Db 128135 CAACACTTGGAGTTTCTGACGCGGATTTGGAGGGGATCGATGAGATTTCCACGACGC 128194

QY 601 CGGCATCAACATTTGGCGAAGGCCACCGATTTCAAAACCCCGGAAATCAACCATCAA 660

Db 128195 CGGCATCAACATTTGGCGAAGGCCACCGATTTCAAAACCCCGGAAATCAACCATCAA 128254

QY 661 CATCAGTTTGTAGTGGCAATTCGGTTCATCACTGCGCAGCAGCTTGTATCCAGCGCA 720

Db 128255 CATCAGTTTGTAGTGGCAATTCGGTTCATCACTGCGCAGCAGCTTGTATCCAGCGCA 128314

QY 721 CACTTGGGGCTGGACAGCGGCGTGAACAATGCTGCTGCGCCGAAACCCACAGCGGGAA 780

Db 128315 CACTTGGGGCTGGACAGCGGCGTGAACAATGCTGCTGCGCCGAAACCCACAGCGGGAA 128374

QY 781 CCAGATCAGCTTTCGCGGAACCGCGCGGGAATTCACCGTTCGGTGTCCCGTA 840

Db 128375 CCAGATCAGCTTTCGCGGAACCGCGCGGGAATTCACCGTTCGGTGTCCCGTA 128434

QY 841 TTGCGCGCGCAGCGCGCGATAAACACAAACCGTTCAAAATACGATTCGGGTTCACCA 900

Db 128435 TTGCGCGCGCAGCGCGCGATAAACACAAACCGTTCAAAATACGATTCGGGTTCACCA 128494

QY 901 GGTGAGCAGATGCTCCATCAACATGGGCTTTACCAAAACCGCGTGTCTATCGAGCTCAC 960
Db 128495 GGTGAGCAGATGCTCCATCAACATGGGCTTTACCAAAACCGCGTGTCTATCGAGCTCAC 128554
QY 961 CTCACCCGACACCGGTTGGCGGTGTGAGTGGCCACCGCGCAACCGCCCAAGCGGTGTC 1020
Db 128555 CTCACCCGACACCGGTTGGCGGTGTGAGTGGCCACCGCGCAACCGCCCAAGCGGTGTC 128614
QY 1021 ATCGGACAGGTTGGTCTGTCTTCTCAATGATCTGGGCGCTTCCACCTGTGTGTGTCAT 1080
Db 128615 ATCGGACAGGTTGGTCTGTCTTCTCAATGATCTGGGCGCTTCCACCTGTGTGTGTCAT 128674
QY 1081 GCGGTCTTTCGCTGCGATGAGCGGAACCATTAACAGTAAGCGATGCGACCCAGCGCAT 1140
Db 128675 GCGGTCTTTCGCTGCGATGAGCGGAACCATTAACAGTAAGCGATGCGACCCAGCGCAT 128734
QY 1141 AATATCGAGCAGATCGCGCGGCTTGGACAAAGATCAACGCCCAAGGTGCGCGGAT 1200
Db 128735 AATATCGAGCAGATCGCGCGGCTTGGACAAAGATCAACGCCCAAGGTGCGCGGAT 128794
QY 1201 GAACAAAGAAGCGTCAGAAATTAACACACGAGAAAGAACCGCAATGATGCTTTCGCGCTT 1260
Db 128795 GAACAAAGAAGCGTCAGAAATTAACACACGAGAAAGAACCGCAATGATGCTTTCGCGCTT 128854
QY 1261 AATCTCTGTTTAAATCACAGTACATTCGGGTCCGATGAGCAAGTAAGATGCGCCCC 1320
Db 128855 AATCTCTGTTTAAATCACAGTACATTCGGGTCCGATGAGCAAGTAAGATGCGCCCC 128914
QY 1321 CAAAGCAGACCTCTAATGAAGATTTCCATGATCACCATCGTAGCTATGAGTACTTA 1380
Db 128915 CAAAGCAGACCTCTAATGAAGATTTCCATGATCACCATCGTAGCTATGAGTACTTA 128974
QY 1381 AGTAAATGATGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATCAACTGGA 1440
Db 128975 AGTAAATGATGTTCTTAACATGTTTAAATATAGCTTCATGAACCCCAATCAACTGGA 129034
QY 1441 CACTTCTCTCATCATGATGAGGAGCTTCGAGGCGCTTCGAGGCGCTTCCAT 1500
Db 129035 CACTTCTCTCATCATGATGAGGAGCTTCGAGGCGCTTCGAGGCGCTTCCAT 129094
QY 1501 TTCCCTCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 1560
Db 129095 TTCCCTCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 129154
QY 1561 GGTATCGCGACCCAAACCGGCAAGCAACCGAGCGGTGAAGTCTTGTGCAAGCAGC 1620
Db 129155 GGTATCGCGACCCAAACCGGCAAGCAACCGAGCGGTGAAGTCTTGTGCAAGCAGC 129214
QY 1621 GCGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCGAGCGCTTGTGA 1680
Db 129215 GCGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCGAGCGCTTGTGA 129274
QY 1681 AATCCCGTTAAACATCGCCATCAACGAGATGCTGCTATCCACATGTTCTCCCGGT 1740
Db 129275 AATCCCGTTAAACATCGCCATCAACGAGATGCTGCTATCCACATGTTCTCCCGGT 129334
QY 1741 CAACAGGTAGCTTCTTGGGTGGAGCAACGCTCACGCTGCGCTTGAAGATGAAGCGCA 1800
Db 129335 CAACAGGTAGCTTCTTGGGTGGAGCAACGCTCACGCTGCGCTTGAAGATGAAGCGCA 129394
QY 1801 CACATTAATCTTGTGCGGTGAGATGTTTGAAGCGGTAAACCGGTGAAGCTAATCC 1860
Db 129395 CACATTAATCTTGTGCGGTGAGATGTTTGAAGCGGTAAACCGGTGAAGCTAATCC 129454
QY 1861 CGTGGCGGATGTGAAGTAGTAGAATCTGGAACCATGCGCACCTTGGCCATGCAACCCC 1920
Db 129455 CGTGGCGGATGTGAAGTAGTAGAATCTGGAACCATGCGCACCTTGGCCATGCAACCCC 129514
QY 1921 CTCATTCGGGATGCTTACATGTTGTGATGGAAACTAGATTGGGCTCGCATGCGCGCTT 1980
Db 129515 CTCATTCGGGATGCTTACATGTTGTGATGGAAACTAGATTGGGCTCGCATGCGCGCTT 129574

QY 1981 AGCTTCGGTCCCAAGATGTCTTCAAGACCGTGACCTGACGGCGCGTCTGATGGTCC 2040
Db 129575 AGCTTCGGTCCCAAGATGTCTTCAAGACCGTGACCTGACGGCGCGTCTGATGGTCC 129634
QY 2041 TGTGGGCGCAGCGCGGTATCCATTTGTCGCGCGGAAGGTTTTTGTGTAGGCAATTCG 2100
Db 129635 TGTGGGCGCAGCGCGGTATCCATTTGTCGCGCGGAAGGTTTTTGTGTAGGCAATTCG 129694
QY 2101 CCGAGGCTTGTGTGGGACCTTCTTCCGAAACCCAGCTGCTCCCATCTCTAAAGCAGG 2160
Db 129695 CCGAGGCTTGTGTGGGACCTTCTTCCGAAACCCAGCTGCTCCCATCTCTAAAGCAGG 129754
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGTATTGGCAACGATGGCG 2220
Db 129755 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGTATTGGCAACGATGGCG 129814
QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGTCTGTGATGACAGCAATCGAGGG 2280
Db 129815 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGTCTGTGATGACAGCAATCGAGGG 129874
QY 2281 ATTGGCGCTTGTGTACTTCTGAAAAGTTTCAAGGTTTTTCACTTCTCGCGCGCAGGAA 2340
Db 129875 ATTGGCGCTTGTGTACTTCTGAAAAGTTTCAAGGTTTTTCACTTCTCGCGCGCAGGAA 129934
QY 2341 TTGGGCGCAGCAGATTAACACCTTTCAGCAATGG 2374
Db 129935 TTGGGCGCAGCAGATTAACACCTTTCAGCAATGG 129968

RESULT 5
AX063767 AX063767 993 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 49 from Patent WO0100843.
DEFINITION AX063767
ACCESSION AX063767
VERSION AX063767.1 GI:12541479
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 993)
AUTHORS Pompejus, M., Krosger, B., Schroeder, H., Zelder, O. and Haberhauer, G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway proteins
JOURNAL Patent: WO 0100843-A 49 04-JAN-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
Location/Qualifiers
source 1..993
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..973
/note="BXA01393"
/codon_start=1
/transl_table=11
/db_xref="GI:12541480"
/protein_id="CAC25125.1"
CDS
233 a 256 c 268 g 236 t
BASE COUNT
ORIGIN
Query Match 41.8%; Score 993; DB 6; Length 993;
Best Local Similarity 100.0%; Pred. No. 2.4e-233;
Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1321 CAAAGCAGACCTCTAATGAAGATTTCCATGATCACCATCGTACCTATGGAAGTACTTA 1380
Db 1 CAAAGCAGACCTCTAATGAAGATTTCCATGATCACCATCGTACCTATGGAAGTACTTA 60

Qy 1381 AGTAAATGATGGTCTTAAATAGCTTTAATATAGCTTCAATGACCCCAATCAACTGGA 1440
Db 61 AGTAAATGATGGTCTTAAATAGCTTTAATATAGCTTCAATGACCCCAATCAACTGGA 120
Qy 1441 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCCTCTTTAGCCCTTTCCAT 1500
Db 121 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCCTCTTTAGCCCTTTCCAT 180
Qy 1501 TTCCCCCTCGCGGTGAGTCAGCGGTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
Db 181 TTCCCCCTCGCGGTGAGTCAGCGGTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 240
Qy 1561 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 1620
Db 241 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 300
Qy 1621 CGGGAATGCTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 1680
Db 301 CGGGAATGCTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 360
Qy 1681 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCATGCTTCCCTCCGCTGT 1740
Db 361 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCATGCTTCCCTCCGCTGT 420
Qy 1741 CAACGAGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTGCGCTTGGAAAGATGAAGCGCA 1800
Db 421 CAACGAGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTGCGCTTGGAAAGATGAAGCGCA 480
Qy 1801 CACATTATCTTGTGCTGCGGTGAGATGTTTAAAGCGGTAAACCCGTGAAGCTAATCC 1860
Db 481 CACATTATCTTGTGCTGCGGTGAGATGTTTAAAGCGGTAAACCCGTGAAGCTAATCC 540
Qy 1861 CGTGGCGGATGCAAGTAGTAGTGGAACTTGGAACTGCGCACTGCGCAATGCGCAACCCC 1920
Db 541 CGTGGCGGATGCAAGTAGTAGTGGAACTTGGAACTGCGCACTGCGCAATGCGCAACCCC 600
Qy 1921 CTCAATCGGGATGCTACATGCTTGAATGGAACTAGATGGGCTGCGATGCCCGCTT 1980
Db 601 CTCAATCGGGATGCTACATGCTTGAATGGAACTAGATGGGCTGCGATGCCCGCTT 660
Qy 1981 ACGCTTGGTCCCAAGATGCTTCAAGACCGTGACCTGGAACCGGCGGTGATGTC 2040
Db 661 ACGCTTGGTCCCAAGATGCTTCAAGACCGTGACCTGGAACCGGCGGTGATGTC 720
Qy 2041 TGTGGGCGGAGCGCTATCATGCTTCCGTCGCGGAGGTTTGGTGAAGCAATTCG 2100
Db 721 TGTGGGCGGAGCGCTATCATGCTTCCGTCGCGGAGGTTTGGTGAAGCAATTCG 780
Qy 2101 CCGAGGCTTGGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 2160
Db 781 CCGAGGCTTGGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 840
Qy 2161 AGAAGTATCTCTCGATGATGATACCATTTGACACACCGATGATTTGGCAACGATGGG 2220
Db 841 AGAAGTATCTCTCGATGATGATACCATTTGACACACCGATGATTTGGCAACGATGGG 900
Qy 2221 CCTGAATCTAGATCTCTAGCTAGACTACAGACCGCTGCTGATGACGCAATCGAGG 2280
Db 901 CCTGAATCTAGATCTCTAGCTAGACTACAGACCGCTGCTGATGACGCAATCGAGG 960
Qy 2281 ATTGCGGCTTGGTACTTCTGAAAGGTTTCAG 2313
Db 961 ATTGCGGCTTGGTACTTCTGAAAGGTTTCAG 993

RESULT 6
AX244055
LOCUS AX244055
DEFINITION Sequence 47 from Patent WO0166573.
ACCESSION AX244055
VERSION AX244055.1 GI:15859119
KEYWORDS linear PAT 29-SEP-2001

SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O., Haberhauer, G.,
Kim, J. W., Lee, H. S. and Hwang, B. J.
TITLE Corynebacterium glutamicum genes encoding metabolic pathway
protein
JOURNAL Patent: WO 0166573-A 47 13-SEP-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
source
1. 993
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101. .973
/note="RXA01393"
/codon_start=1
/transl_table=11
/protein_id="CAC88506.1"
/db_xref="GI:15859120"
translation="MNPQLDLSIIIDGSEFGASLALSISPSAVSRVQKALEHHVG
RVLSRTQPAKATEAGELVQAAKRWLLQAEKALSGRLAEIPLTAINADLSLSTW
FPFVNEVASWGGATLRLDEAHTLSLRGVDLGAATREANVAGCEVVELGTM
HLAIATPSLRDAYMVDGLDWAAMPVLRFGKVDLQDRDLDRVDGPGVRRVSVPS
AEGFGEAIRRGLGWGLLPETQNAFMLKAGEVILLDEIPIDFPMYQWRWLSRSRLRL
TDVVDDAAIEGLRP"
BASE COUNT 233 a 256 c 268 g 236 t
Query Match 41.8%; Score 993; DB 6; Length 993;
Best Local Similarity 100.0%; Pred. No. 2.4e-233; Indels 0; Gaps 0;
Matches 993; Conservative 0; Mismatches 0;
Qy 1321 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
Db 1 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 60
Qy 1381 AGTAAATGATGGTCTTAAACATGTTTAAATATAGCTTCAATGACCCCAATCAACTGGA 1440
Db 61 AGTAAATGATGGTCTTAAACATGTTTAAATATAGCTTCAATGACCCCAATCAACTGGA 120
Qy 1441 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCCTCTTTAGCCCTTTCCAT 1500
Db 121 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCCTCTTTAGCCCTTTCCAT 180
Qy 1501 TTCCCCCTCGCGGTGAGTCAGCGGTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
Db 181 TTCCCCCTCGCGGTGAGTCAGCGGTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 240
Qy 1561 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 1620
Db 241 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 300
Qy 1621 CGGGAATGCTGTTGCTGCAAGCAGAACTAAAGCGGTAAACCCGTGAAGCTTGGCTGA 1680
Db 301 CGGGAATGCTGTTGCTGCAAGCAGAACTAAAGCGGTAAACCCGTGAAGCTTGGCTGA 360
Qy 1681 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCATGCTTCCCTCCGCTGT 1740
Db 361 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCATGCTTCCCTCCGCTGT 420
Qy 1741 CAACGAGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTGCGCTTGGAAAGATGAAGCGCA 1800
Db 421 CAACGAGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTGCGCTTGGAAAGATGAAGCGCA 480
Qy 1801 CACATTATCTTGTGCTGCGGTGAGATGTTTAAAGCGGTAAACCCGTGAAGCTAATCC 1860
Db 481 CACATTATCTTGTGCTGCGGTGAGATGTTTAAAGCGGTAAACCCGTGAAGCTAATCC 540
Qy 1861 CGTGGCGGATGCAAGTAGTAGTGGAACTTGGAACTGCGCACTGCGCAATGCGCAACCCC 1920


```

Db 541 CGTGGCGGATGTGAAGTAGTAGAATTTGGAACCATCGCCACTTGGCCATTGCAACCCC 600
QY 1921 CTCAATTCGGGATGCCCTACATGTTGATGGAATACTAGATTGGCTCGGATGCCGCTT 1980
Db 601 CTCAATTCGGGATGCCCTACATGTTGATGGAATACTAGATTGGCTCGGATGCCGCTT 660
QY 1981 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTAGCTGACCGGCGCGCTCGATGCTCC 2040
Db 661 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTAGCTGACCGGCGCGCTCGATGCTCC 720
QY 2041 TGTGGGCGCAGGCGCTATTCATTTGCTCCGTCGGCGAAGTTTGGTCAAGCAATTG 2100
Db 721 TGTGGGCGCAGGCGCTATTCATTTGCTCCGTCGGCGAAGTTTGGTCAAGCAATTG 780
QY 2101 CCGAGGCGCTTGGTGGGACCTTCTCCGGAACCAAGCTGCTCCCATGTAAAGCAGG 2160
Db 781 CCGAGGCGCTTGGTGGGACCTTCTCCGGAACCAAGCTGCTCCCATGTAAAGCAGG 840
QY 2161 AGAAGTCACTCTCTCGATGATACCATTTGACACACCGCATTTATTTGCAAGTGGCG 2220
Db 841 AGAAGTCACTCTCTCGATGATACCATTTGACACACCGCATTTATTTGCAAGTGGCG 900
QY 2221 CCTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGTGGTGTGATCAGCAATCGAGG 2280
Db 901 CCTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGTGGTGTGATCAGCAATCGAGG 960
QY 2281 ATTGGCGCTTAGTACTTCTGAAAGGTTTCAG 2313
Db 961 ATTGGCGCTTAGTACTTCTGAAAGGTTTCAG 993

RESULT 7
AX123540
LOCUS AX123540 870 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3456 from Patent EP1108790.
ACCESSION AX123540
VERSION AX123540.1 GI:14041028
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 870)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3456 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source
1. .870
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 192 a 236 c 245 g 197 t
ORIGIN

Query Match 36.6%; Score 870; DB 6; Length 870;
Best Local Similarity 100.0%; Pred. No. 4.4e-203;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1421 ATGAACCCCAATCACTGGACACTTGTCTCAATCAATGATGAAGCAGCTTCGAAGGC 1480
Db 1 ATGAACCCCAATCACTGGACACTTGTCTCAATCAATGATGAAGCAGCTTCGAAGGC 60
QY 1481 GCCTCTTAGCCCTTTCCATTTCCCTTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 1540
Db 61 GCCTCTTAGCCCTTTCCATTTCCCTTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 120
QY 1541 CATCACTGGGTGAGTGTGGTATCGGCAACCAACCGCGCAAGCAACCGAAGCGGT 1600
Db 121 GATCACTGGGTGAGTGTGGTATCGGCAACCAACCGCGCAAGCAACCGAAGCGGT 180
QY 1601 GAAGTCTTGTGCAAGCAGCGCGGAAAATGTTGTCTGCAAGCAGAACTAAAGCGCAA 1660

```

```

Db 181 GAAGTCTTGTGCAAGCAGCGCGGAAAATGTTGTCTGCAAGCAGAACTAAAGCGCAA 240
QY 1661 CTATCTGGAGCGCTTGTCTGAAATCCCGTTAACCATCGCCATCAACGACAGATTGCTATCC 1720
Db 241 CTATCTGGAGCGCTTGTCTGAAATCCCGTTAACCATCGCCATCAACGACAGATTGCTATCC 300
QY 1721 ACATGTTTCTCCCGCTGTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTG 1780
Db 301 ACATGTTTCTCCCGCTGTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTG 360
QY 1781 CGCTTGGAGATGAGCGCACACATTTCTTCTGCGCGCTGGAGATGTTTATGAGAGCG 1840
Db 361 CGCTTGGAGATGAGCGCACACATTTCTTCTGCGCGCTGGAGATGTTTATGAGAGCG 420
QY 1841 GTAAACCCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAACTTGGAAACCATGCGC 1900
Db 421 GTAAACCCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAACTTGGAAACCATGCGC 480
QY 1901 CACTTGGCCATTGCAACCCCTCATTTGCGGGATGCCATCATGTTGATGGGAAACTAGAT 1960
Db 481 CACTTGGCCATTGCAACCCCTCATTTGCGGGATGCCATCATGTTGATGGGAAACTAGAT 540
QY 1961 TGGGCTGCGATGCGCGTCTTACGCTTCCGTCCTCAAGATGTCTTCAAGACCGTACCTG 2020
Db 541 TGGGCTGCGATGCGCGTCTTACGCTTCCGTCCTCAAGATGTCTTCAAGACCGTACCTG 600
QY 2021 GACGGCGCTGATGCTCTGTTGGGCGCAGCGCTATTCATTGTCCTCGCGGAA 2080
Db 601 GACGGCGCTGATGCTCTGTTGGGCGCAGCGCTATTCATTGTCCTCGCGGAA 660
QY 2081 GGTGTTGTGAGCAATTCGCGAGCGCTTGTGTTGGGAGCTTCTTCCGAAACCCAGCT 2140
Db 661 GGTGTTGTGAGCAATTCGCGAGCGCTTGTGTTGGGAGCTTCTTCCGAAACCCAGCT 720
QY 2141 GCTCCCATCTAAAGCAGGAGATGTCTCTCGATGAGATATCCATTGACACACCG 2200
Db 721 GCTCCCATCTAAAGCAGGAGATGTCTCTCGATGAGATATCCATTGACACACCG 780
QY 2201 ATGATTTGCAACGATGCGCGCTTGAATCTAGATCTCTAGTAGACTACAGACCGCTC 2260
Db 781 ATGATTTGCAACGATGCGCGCTTGAATCTAGATCTCTAGTAGACTACAGACCGCTC 840
QY 2261 GTTGTGAGCAATTCGAGGATTTGGGCT 2290
Db 841 GTTGTGAGCAATTCGAGGATTTGGGCT 870

RESULT 8
AX063771/c
LOCUS AX063771 822 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 53 from Patent WO0100843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 822)
AUTHORS Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
TITLE corynebacterium glutamicum genes encoding metabolic pathway
JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;
BASF AKTIENGESELLSCHAFT (DE)
FEATURES
source
1. .822
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101. .802
/note="RXA01394"
/codon_start=1
CDS

```



```
/transl_table=11
/protein_id="CAC25127.1"
/db_xref="GI:12541484"
/translation="MEIPITGLLGLSLLSIPONVLIKQIKREGIALLVCLL
SDVFLIAGTLDVLLSNAAPVLDIMRWGGIAYLLWFAVMAKADMTNKVAPQIIE
ETPTVPDDTPLGGSVAVIDTRNRVRVSVVKORVWVKPMLMAIIVLTWLNPNAYLDA
FVPIGVGAQYDGTGRWIFPAAGAFASLIWFLVPGFGAALSRLPSLPKVRWVNVV
AVMTALAIAKLMLMG"
BASE COUNT 167 a 192 c 246 g 217 t
ORIGIN

Query Match 34.6%; Score 822; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.9e-191;
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCAAAACCCGCGAAACTTAACCATCAACATCAGTTTGTATGCGCAATGCGGTCAATC 688
DB 822 GATTCAAAACCCGCGAAACTTAACCATCAACATCAGTTTGTATGCGCAATGCGGTCAATC 763
QY 689 ACAATGCCACGACGAGCTTGTATCCAGCGCCACACTTGGGGCTGGACAGCGGGCGGTGAC 748
DB 762 ACAATGCCACGACGAGCTTGTATCCAGCGCCACACTTGGGGCTGGACAGCGGGCGGTGAC 703
QY 749 AATGCTCTGCGCGAAACCCACAGCGGGAACAGATCAGGCTTGGCGGAAACGCGCCA 808
DB 702 AATGCTCTGCGCGAAACCCACAGCGGGAACAGATCAGGCTTGGCGGAAACGCGCCA 643
QY 809 GCGCGAAATCCACCTGCGGTGTGCGCGTATTTGGCGCGGACCGCGCGATTAACACA 868
DB 642 GCGCGAAATCCACCTGCGGTGTGCGCGTATTTGGCGCGGACCGCGCGATTAACACA 593
QY 869 AACGCGTCCAAATACGATTCGCGGTTCACACAGCTCAGCACGATTTGCCATCAACATGGGC 928
DB 582 AACGCGTCCAAATACGATTCGCGGTTCACACAGCTCAGCACGATTTGCCATCAACATGGGC 523
QY 929 TTTACCCAAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 988
DB 522 TTTACCCAAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 463
QY 989 GTGGCCACCGCGGAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 1048
DB 462 GTGGCCACCGCGGAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 403
QY 1049 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGCTTTTCGCTGCCATGACGCGCAAC 1108
DB 402 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGCTTTTCGCTGCCATGACGCGCAAC 343
QY 1109 CATAACAGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTTG 1168
DB 342 CATAACAGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTTG 283
QY 1169 GACAAAAGATCAACGCCAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 1228
DB 282 GACAAAAGATCAACGCCAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 223
QY 1229 ACCGAGAAACCCGATGAGTCTTTCGCGTAAATTCCTTTTAAATCACCAGTACATTC 1288
DB 222 ACCGAGAAACCCGATGAGTCTTTCGCGTAAATTCCTTTTAAATCACCAGTACATTC 163
QY 1289 TGGGTCGATGACAGTAAAGACTGGCCCCCAAGCAGACCTGTAAATGAAGATTTC 1348
DB 162 TGGGTCGATGACAGTAAAGACTGGCCCCCAAGCAGACCTGTAAATGAAGATTTC 103
QY 1349 ATGATCACCATCGTGAACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 1408
DB 102 ATGATCACCATCGTGAACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 43
QY 1409 TAATATAGCTTTCATGAACCCATTCACCTGGACACTTTGCTC 1450
DB 42 TAATATAGCTTTCATGAACCCATTCACCTGGACACTTTGCTC 1
```

RESULT 9

```
AX244059/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 822)
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O., Haberhauer, G.,
Kim, J. W., Lee, H. S. and Hwang, B. J.
Corynebacterium glutamicum genes encoding metabolic pathway
proteins
Patent: WO 0166573-A 51 13-SEP-2001;
BASF AKTIENGESSELLSCHAFT (DE)
FEATURES
Location/Qualifiers
1..822
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..802
/note="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC8508.1"
/db_xref="GI:15859124"
/translation="MEIPITGLLGLSLLSIPONVLIKQIKREGIALLVCLL
SDVFLIAGTLDVLLSNAAPVLDIMRWGGIAYLLWFAVMAKADMTNKVAPQIIE
ETPTVPDDTPLGGSVAVIDTRNRVRVSVVKORVWVKPMLMAIIVLTWLNPNAYLDA
FVPIGVGAQYDGTGRWIFPAAGAFASLIWFLVPGFGAALSRLPSLPKVRWVNVV
AVMTALAIAKLMLMG"
BASE COUNT 167 a 192 c 246 g 217 t
ORIGIN

Query Match 34.6%; Score 822; DB 6; Length 822;
Best Local Similarity 100.0%; Pred. No. 2.9e-191;
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCAAAACCCGCGAAACTTAACCATCAACATCAGTTTGTATGCGCAATGCGGTCAATC 688
DB 822 GATTCAAAACCCGCGAAACTTAACCATCAACATCAGTTTGTATGCGCAATGCGGTCAATC 763
QY 689 ACAATGCCACGACGAGCTTGTATCCAGCGCCACACTTGGGGCTGGACAGCGGGCGGTGAC 748
DB 762 ACAATGCCACGACGAGCTTGTATCCAGCGCCACACTTGGGGCTGGACAGCGGGCGGTGAC 703
QY 749 AATGCTCTGCGCGAAACCCACAGCGGGAACAGATCAGGCTTGGCGGAAACGCGCCA 808
DB 702 AATGCTCTGCGCGAAACCCACAGCGGGAACAGATCAGGCTTGGCGGAAACGCGCCA 643
QY 809 GCGCGAAATCCACCTGCGGTGTGCGCGTATTTGGCGCGGACCGCGCGATTAACACA 868
DB 642 GCGCGAAATCCACCTGCGGTGTGCGCGTATTTGGCGCGGACCGCGCGATTAACACA 583
QY 869 AACGCGTCCAAATACGATTCGCGGTTCACACAGCTCAGCACGATTTGCCATCAACATGGGC 928
DB 582 AACGCGTCCAAATACGATTCGCGGTTCACACAGCTCAGCACGATTTGCCATCAACATGGGC 523
QY 929 TTTACCCAAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 988
DB 522 TTTACCCAAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 463
QY 989 GTGGCCACCGCGGAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 1048
DB 462 GTGGCCACCGCGGAACCCGCGTGTATCGACGCTCACCTCCACCGCGACCGCGCGATTAACACA 403
QY 1049 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGCTTTTCGCTGCCATGACGCGCAAC 1108
DB 402 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGCTTTTCGCTGCCATGACGCGCAAC 343
QY 1109 CATAACAGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTTG 1168
DB 342 CATAACAGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTTG 283
QY 1169 GACAAAAGATCAACGCCAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 1228
DB 282 GACAAAAGATCAACGCCAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 223
QY 1229 ACCGAGAAACCCGATGAGTCTTTCGCGTAAATTCCTTTTAAATCACCAGTACATTC 1288
DB 222 ACCGAGAAACCCGATGAGTCTTTCGCGTAAATTCCTTTTAAATCACCAGTACATTC 163
QY 1289 TGGGTCGATGACAGTAAAGACTGGCCCCCAAGCAGACCTGTAAATGAAGATTTC 1348
DB 162 TGGGTCGATGACAGTAAAGACTGGCCCCCAAGCAGACCTGTAAATGAAGATTTC 103
QY 1349 ATGATCACCATCGTGAACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 1408
DB 102 ATGATCACCATCGTGAACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 43
QY 1409 TAATATAGCTTTCATGAACCCATTCACCTGGACACTTTGCTC 1450
DB 42 TAATATAGCTTTCATGAACCCATTCACCTGGACACTTTGCTC 1
```

```

Db 342 CATTAAGGTAAGCGATGCCACCCAGCGCATATATCGAGACGATCGCGCGGCGATTG 283
QY 1169 GACAAAAGATCAAGCGCCAAAGGTGCGCGCGATGAACAAAAGACGTGAGAAATTAACAC 1228
Db 282 GACAAAAGATCAAGCGCCAAAGGTGCGCGCGATGAACAAAAGACGTGAGAAATTAACAC 223
QY 1229 ACAGAGAAGACCGCAATGAGTCTTTCGCGCTTAATCTCTTGTGTTAATCAGCAGTACATTC 1288
Db 222 ACAGAGAAGACCGCAATGAGTCTTTCGCGCTTAATCTCTTGTGTTAATCAGCAGTACATTC 163
QY 1289 TCGCGTCCGATGACAGTAAAGACTGCGCCCAAGACGAGCTGTAATGAAGATTTC 1348
Db 162 TCGCGTCCGATGACAGTAAAGACTGCGCCCAAGACGAGCTGTAATGAAGATTTC 103
QY 1349 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATGTTGTTCTTAACATGGTT 1408
Db 102 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATGTTGTTCTTAACATGGTT 43
QY 1409 TAATATAGCTTCAAGAACCCCATTCACACTGGACACTTTGCTC 1450
Db 42 TAATATAGCTTCAAGAACCCCATTCACACTGGACACTTTGCTC 1

RESULT 10
AB083133/c
LOCUS
DEFINITION
Corynebacterium efficiens lysG, lysE genes for lysine export
transcriptional regulatory protein, Lysine exporter protein,
complete cds.
ACCESSION
AB083133
VERSION
AB083133.1 GI:20065738
KEYWORDS
Corynebacterium efficiens DNA.
SOURCE
Corynebacterium efficiens
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterinae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
lysG, lysE of Corynebacterium efficiens
Published Only in Database (2002)
2 (bases 1 to 1771)
Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
Direct Submission
Submitted (04-APR-2002) Hiroshi Itaya, AJINOMOTO CO., INC,
Fermentation & Biotechnology Laboratories; 1-1, Suzuki-Cho,
Kawasaki-Ku, Kawasaki, Kanagawa 210-8681, Japan
(E-mail: hitoshi.itaya@ajinomoto.com, Tel: 81-44-244-7123 (ex. 4146),
Fax: 81-44-222-0129)
FEATURES
Source
1. 1771
/organism="Corynebacterium efficiens"
/db_xref="taxon:152794"
complement(47..919)
/gene="lysG"
complement(47..919)
/gene="lysG"
/codon_start=1
/transl_table=1
/product="Lysine export transcriptional regulatory
protein"
/protein_id="BAB88826.1"
/db_xref="GI:20065739"
/transl_table="NNPIHDLTLLTIIDGSPENASLALSISPSAVSORIKALEKSVG
RVLVSPTQPAVEAGEVLVQARLQAEATREOLAEALDEIPLTVAINADSLSTW
PPVPAEVAHWGAVTLTLVEDEAHTLSLLRGSVLGAVTREADPVAGCEVLRIGVMR
HLFVATPELARVTVDGQDVMPLVREGPNVLDLDRLEGRVDGAVARRSVVPS
AESGFAVRILGNGLLPEQAAPMLAAGDVQLDEKVDVDTFLYQWRKLESRLRLARL
TDAVDAARAGLRT"
981..1676
/gene="lysE"
981..1676
/gene="lysE"

```

```

/codon_start=1
/transl_table=1
/product="Lysine exporter protein"
/protein_id="BAB88827.1"
/db_xref="GI:20065740"
/transl_table="WRDMEIFVTGLLIGASLLAIGPONLVIKOGIKREGITAVIIV
CLLSDVLFITGLTGLVGLISDTAPIILDILRWCGIAYLLWFAWAAARDALRARTVTF
VEHSEPVAAASAGSGVTTKORPLRITSGTQVWVWPMMLMAI VLTWLPNAYLLDAFV
FIGVGAYQYGETGRMIFAAAGAFASLVWFLVGVGYAAALSRLSPRYWRWVINITGAV
VLTGLAVKLIILMG"
BASE COUNT 294 a 603 c 545 g 329 t
ORIGIN
Query Match 33.6%; Score 798.8; DB 1; Length 1771;
Best Local Similarity 68.4%; Pred. No. 1.6e-185;
Matches 1158; Conservative 0; Mismatches 517; Indels 19; Gaps 3;
QY 650 TAACCCATCAACATCAGTTTGTATGCGCAATGCGGTTCATCACAACTGCCACGACGAGTTG 709
Db 1675 TAACCCATCAGGATCAGTTTACGGCCCAATCCGGTGAGCACCGGCCACACCTATGTTG 1616
QY 710 ATCCAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGAACAATGCTGCTGCGCGCAACCC 769
Db 1615 ATCCAGCGCCACCGCGGGCTGGACAGGGGACGCGACAGTGGCGCGCGGTAAACCG 1556
QY 770 ACAGCGGGAACAGATCAGGCTTGGCGGAACGCGGCCAGCGGCGGGAATATCACCGTCCG 829
Db 1555 ACCAGTGGGAACAGACAGGCTGGCGGCGAAGGACCGCGAGCGAAGATCCACCGACCG 1496
QY 830 GTGTGCGCGTATTTGGCGCGGCGGCGGATAAACAAACGCGTCCAAATACGATTC 889
Db 1495 GTCTCCCGTACTGGGCTCCGACACCGCGATGAAGACGAAGGCATCCAGGTAGGCATTG 1436
QY 890 GGGTTCAACAGGTGACGACAGTTCATCAATGCGGTTCATCCAAACCCGCTGCTTA 949
Db 1435 GGATTCAGCCAGGTGACGACATATGCCATGAGCATGGGCTGACCCAGACCTTCGCGGTG 1376
QY 950 TCGAGCGCTCACCTCCACCGCACCGCGTTGCGGTGTCAGTGGCCACCGCCGCAACCGCCC 1009
Db 1375 CTTGATGTGATCGGAGCGCGGTGCTGTTTTCGTCGTCACGCCCGCGCGAGGCGGAC 1316
QY 1010 AAGGCGTGTCTATCGGCGACGTTGTTCTGTTCTTCAATGATCTGTGCGGCTTCCACC 1069
Db 1315 GCTGCGG---CAACGGTTCGGAATGCTC-----GACAAAGTTTACTCGGTG 1271
QY 1070 TTGTTTGTATGCGGTCTTTTCGTCGCATGACGCGCAACCAATACAGGTAAAGCGTCCA 1129
Db 1270 CGGGCGCGCAGGGCGTTCGCGCGCGCCATCATCCGCGAACCCACAGCAGGTAGGCGATG 1211
QY 1130 CCCAGCGCATTAATATCGAGCAGATCGCGCGGCGATTCGACAAAAGATCAAGCCCAAG 1189
Db 1210 CACACGCGAGGATGTGAGAAATGATCGGGCGGTGTCGAGATCAGGCGCGACCCCGAGG 1151
QY 1190 GTGCGGCGATGAACAAAAGACGTGAGAAATTAACACACAGAGAGACCCCAATGCT 1249
Db 1150 GTGCGGAGGTGAACAGGACCGACGTCGCGACAGCAGACAGATGATGACGCGCGTGTG 1091
QY 1250 CTTTCGCGCTTAATCTCTTGTGTTTAAATCACCAGTACATTCCTGCGGTCCGATGACAGTAA 1309
Db 1090 CCTCGCGTTTGTGTCCTGTTTGTATCACCAGGACATCTCTGTGGCGCGATGCCAACAGC 1031
QY 1310 AGACTGGCCCCCAAGACGAGCTGTAATGAAGATTTCCATGATCAACCATCGTGTGACTAT 1369
Db 1030 AGACTGGCTCCCAACAAACAAACCGGTGAACAAAGATTTCCATGTCCCGCATGTTTCGCGC 971
QY 1370 GGNAGTACTTAAGTAAATGATTGTTCTTAACATGTTTTTAATATAGCTTCATGACCC 1429
Db 970 AAACCACTTAAGTAAATGATGAGACTGAACATGTTTAGAATTCCTTCATGAACCCG 911
QY 1430 ATTCAACTGGACACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAAAGCGCGCTCTTA 1489
Db 910 ATCCACCTGGACACCCCTCTCTGACCATCATCGAGGAGGCGAGCTTCGAGAACCGCTCCCTC 851

```

QY 1490 GCCTTTCCATTTCCCTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCAAGTG 1549
Db 850 GCATGTGCGATCTCCCTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCAAGTG 791
QY 1550 GGTGCGAGTGTGATCGCGACCCCAACCGGCCCAACCGAGCGGTGAGTCCCTT 1609
Db 790 GGTGCGGTACTGTGTCGCGACCCCAACCGGCCCAACCGAGCGGTGAGTCCCTT 731
QY 1610 GTCAAGCAGCGCGGGAATGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGA 1669
Db 730 GTCCAGCGCGCGCAGATGCGCTCTGAGCGCGGAGACCGTGAACACTGCGGAA 671
QY 1670 CGCTTGTGAATCCCGTTAAACATGCGCATCAACGAGATTCGCTATCCATGTTT 1729
Db 670 CGCTTTGACGAGATCCCTCGACCGTGGCTATCAACCGCGATTCCTGTCCATGTTT 611
QY 1730 CTTCCCGTGTCAACGAGGTAGCTCTTGGGTGGAGCAAGCTCAAGTGGCTGGAA 1789
Db 610 CCGCGCTGTTCGCGAGGTGACATTTGGGTGGAGCAAGCTCAAGTGGCTGGAA 551
QY 1790 GATGAAGCGCACATTTATCTTCTGCGCGCTGGAGATGTTTTAGAGCGGTAAACCGT 1849
Db 550 GACGAGCGCCACACCTGTCTCTGCGCGCGGTTCGCTGCGGGCGGTGACCGCG 491
QY 1850 GAAGCTAATCCCGTGGCGGATGTAGTAGTAGTAAGATTTGGAAACCATGCGCCACTTGCC 1909
Db 490 GAGCGGACCCCGTTGCGCGGTGTGAGTCTACCGCTCGGGGTGATGAGACACCTGCCG 431
QY 1910 ATTGAACCCCTCATTTGCGGGATGCTACATGTTGATGGGAACTAGATTTGGCTGG 1969
Db 430 GTGGCACCCCGGACTGAGGGCAAGGTACACCGTGGATGGCCAAACCGGACTGGGTGGC 371
QY 1970 ATGCGCGCTTTAGCGTTCGCTCCCAAGATGTCTTCAAGACCGTGCACCTGGAGCGCGC 2029
Db 370 ATGCGCGTACTGCTTTCGCGCGGATGTGCTGAGGACCGCATCTCGAGGCGAG 311
QY 2030 GTGATGTCCTGTGGGCGCAGCGCGTATCCATTTGCTGCGCGGAGGTTTTGGT 2089
Db 310 GTGCGAGGAGCGCTGCGCAGCGCGCTGCTGCGGTGCTGCGGAGGTTTTGGG 251
QY 2090 GAGGCAATTCGCGAGCGCTTGTGGGACTTCTTCCGAAACCGAGTGTCTCCGATG 2149
Db 250 GAGCGGTGCGACTGCGCGCTGGGGTGGGGCTGCTGCGGAGCGCGCGCGATG 191
QY 2150 CTAAAGCAGGAGAGTGTCTCTCGATGAGATACCCATTCACACCGCATGTTGG 2209
Db 190 CTGGCTCGCGGATGTGCTGAGTGTGATGAGAGGTGTGACACCGCATGTTGG 131
QY 2210 CAACGATGCGCGCTGGAATCTAGATCTAGCTAGTACTACAGACCGCGTGTGATGCA 2269
Db 130 CAACGATGCGCGCTGAGTCCAGGCTGCTGCGCAGACTCACAGACCGCGTGTGATGCG 71
QY 2270 GCAATCGAGGATGCGCGCTTGTAGTACTTCTGAAAGTTGAGGTTTTCTTCTTTC 2329
Db 70 GCGCGCGGCTGTAG---AACCTAAACCTGGAACGTTCCGGTCTTGTACCTCTTC 15
QY 2330 GCGCGAGGAATTG 2343
Db 14 ACCGAGGATATCG 1

RESULT 11
AX123539/c
LOCUS AX123539
DEFINITION Sequence 3455 from Patent EP1108790.
ACCESSION AX123539
VERSION AX123539.1 GI:14041027
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.

REFERENCE 1 (bases 1 to 708)
AUTHORS Nakagawa, S., Mizouchi, H., Ando, S., Hayashi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
Location/Qualifiers
1..708
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718" 180 t
BASE COUNT 134 a 173 c 221 g 180 t
ORIGIN
Query Match 29.8%; Score 708; DB 6; Length 708;
Best Local Similarity 100.0%; Pred. No. 3.3e-163;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 652 ACCATCAACATCAGTTTGTATGCGCAATGCGTCTATCAAACTGCGACGACGTTGAT 711
Db 708 ACCATCAACATCAGTTTGTATGCGCAATGCGTCTATCAAACTGCGACGACGTTGAT 649
QY 712 CCAGCGCACACCTTGGGGCTGAGCAGCGGGCTGACATGCTGTGGCGCAACCCAC 771
Db 648 CCAGCGCACACCTTGGGGCTGAGCAGCGGGCTGACATGCTGTGGCGCAACCCAC 589
QY 772 CAGCGGGAACCAAGATCAGGCTTTCGCGCAACCGCGCAGCGGCAAAATCCACCGTCCGT 831
Db 588 CAGCGGGAACCAAGATCAGGCTTTCGCGCAACCGCGCAGCGGCAAAATCCACCGTCCGT 529
QY 832 GTCGCGTATTGCGCGCAGACCGCGCGATAAACAAACGCGTCCAAATACGATTCGG 891
Db 528 GTCGCGTATTGCGCGCAGACCGCGCGATAAACAAACGCGTCCAAATACGATTCGG 469
QY 892 GTTCAACAGGTGAGCAGATTTGCCATCAAATGGGCTTTACCCAAACCGCGTCTTATC 951
Db 468 GTTCAACAGGTGAGCAGATTTGCCATCAAATGGGCTTTACCCAAACCGCGTCTTATC 409
QY 952 GACGCTCACTCCACCGCACCGCGTTCGCGTGTGAGTGGCCACCGCGCAACCCCAA 1011
Db 408 GACGCTCACTCCACCGCACCGCGTTCGCGTGTGAGTGGCCACCGCGCAACCCCAA 349
QY 1012 AGCGGTGTCATCGGCGCAGGTTGTTCTTCTTCAATGATCTGTGGCGTTCCACCTT 1071
Db 348 AGCGGTGTCATCGGCGCAGGTTGTTCTTCTTCAATGATCTGTGGCGTTCCACCTT 289
QY 1072 GTTGTGATGCGCTTTCGCTGCCATGACGCGCAACCATTAACAGGTAGCGATGCCACC 1131
Db 288 GTTGTGATGCGCTTTCGCTGCCATGACGCGCAACCATTAACAGGTAGCGATGCCACC 229
QY 1132 CCAGCGCATATATCGAGCAGCATCGCGCGCATTTGACAAAAGATCAACGCCCAAGGT 1191
Db 228 CCAGCGCATATATCGAGCAGCATCGCGCGCATTTGACAAAAGATCAACGCCCAAGGT 169
QY 1192 GCCGCGCATGAACAAAAGAGCTCAGAAATTAACACACGAGAAAGACCGCAATGATGCC 1251
Db 168 GCCGCGCATGAACAAAAGAGCTCAGAAATTAACACACGAGAAAGACCGCAATGATGCC 109
QY 1252 TTCGCGCTTAATTCCTTGTATTAATCAGCATACATTCGCGTCCGATGGACAGTAAAG 1311
Db 108 TTCGCGCTTAATTCCTTGTATTAATCAGCATACATTCGCGTCCGATGGACAGTAAAG 49
QY 1312 ACTGCGCGCGCAACAGCAGCTGTAAATGAATTTCCATGATCACCAT 1359
Db 48 ACTGCGCGCGCAACAGCAGCTGTAAATGAATTTCCATGATCACCAT 1
RESULT 12
AX123538
LOCUS AX123538
DEFINITION Sequence 3454 from Patent EP1108790.
ACCESSION AX123538
VERSION AX123538.1 GI:14041026
KEYWORDS

[illegible]

```
/translation="MPRKLDLRTGRPVWSAYRAPAVPTSALTRDVTVDVLIIGMGISG
```


HGAYDTHLMDPANAKAMAQAIETALIADAGNAATYQANTKKLIDDLALDAEVVET
 VKFVKPFTVFHDAYOFYFHRYSKVTAGSITVSPETLPGADRVKQMOBKVROLGAPC
 VFAEPQPEKPLSVITETGTAASKATLDPPEATLTPGPDLYFKLMRGISAGSKNCLS"
 complement (1099...2166)
 /gene="AGR_C_2807"
 complement (1099...2166)
 /note="TRANSCRIPTION REGULATOR AGLR"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_2807p"
 /protein_id="AAK87313.1"
 /db_xref="GI:15156610"
 /translation="MNRNDMDTKSGGSDMPGAMGRPTLKTIAVMTGLITVSGRA
 LKADPDIGASTKERVRLIARQIGVQPNRAGVRLTGKTNVIALVLSVDELMGFTSQM
 VFGITVLSITQVHLVTPHIAKDSMPVIRYILETGSADGVLIISKLENDPRVRMT
 ERNPFVTHGRSDMGIEHAFHDFDNEAYEAVERLAOCQKRIAVIPPSRPFHDH
 ARGFNRGIDFGITFPIDAVITETPLEKIRDFQRLMQSSDRPDGIVSISGSTIA
 LVAGFEAGVKGIGEDVDIVSKQSAEFLNWKIPQIHTVNEIDIKLAGRELAKALLARIG
 EAAETLQSIGSPVSSMAPKP"
 2640...5138
 /gene="AGR_C_2809"
 2640...5138
 /note="hypoetical protein"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_2809p"
 /protein_id="AAK87314.1"
 /db_xref="GI:15156611"
 /translation="MQGPMISSSTPTETVIDLAGLWLASVEGDHATEISIPGDIHSA
 LKNAAIIPQYHGANEKAVOWAQODWIIERTFILDARASWYLDIDYLDVIAIVFN
 DVLVADNCFRVRPDISRAVPGENTLRIHPHSNITAGARQARQPPVYPHPGNS
 PIANGMLKPKQCHFGWDRNIAIAPLGILVKILLKRLDRIARIEHVSSQHVGGVEL
 HVAUTLFAEGPASPILVLSGLDERLKECGVAGETVVRHVFPEVNDLWVWAGSGEQ
 TLYKLTPELDETFTQIGRTIELTDXDEAGSFRAFRINGREIFRCRANMIPADNL
 YLSKTEBEDLCSAVEANMMNIRVWGGGFYEDWYFDLCRLGLLVWQDFPACNL
 YPCSDPFLONVEHYOVKRLSHSPSIALWCGDNELVGLTWFDERSNNRDRYLVA
 DRLNRTIELKALKEATPEALWPPSPASGYLDYGDADHAGSDGMHVMYSWHENKSPDN
 YHOKPRFCSEGFQSYTSMPTVITYAEDKDNMNIASPVIELHQVNGNERIAGTNFR
 YRFPDFENFVLSQVQKALRTAVDWRSKPHCMGTLVWQNDTWPFVSWSSLD
 YGGKXALHYAFOPVAVASAIPSADGRVTFVSNVDNDAEDVEIDMNIVALMDGN
 RVLKANGTCTSDKAATLTDIMDSLPDGAILAWNFIASNGMTGEGHVRDITYKALE
 LQPLAGLEFVGLPKNGQFIEDVTAAGLALFMLEADQPGKRSYDNLFDLAAGETRIIP
 TPKAGPQPHFRIFDLHTCQSSPNPGIETMRKA"
 5278...6273
 /gene="AGR_C_2811"
 5278...6273
 /note="hypoetical protein b3001 - Escherichia coli
 (strain K-12)"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_2811p"
 /protein_id="AAK87315.1"
 /db_xref="GI:15156612"
 /translation="MKYNHCGKTLGLPAISLGLWHNFNDTPHQTKQALCRRAPDLG
 ITHFDLANNYPGPPGSAETAFAEILKTDPRGVRDEMLISKAGINWPGFYGEWSRK
 YLISCCQSLKRMGLDYDIFYSHRFPDNPTEETGALDQIVRSGKALYVGSISYNS
 KTRTEAAALKDGTGTCIIHOPYSYMINRIBEDGLVDLTELIGSIVFSLAQML
 TTKYLGVPDGGRSQSKSLNPAFLNERNVENIRALNSIABRRGOTLAQMAIAWVLRG
 GRITSALIGASRVQVEDCVGLDNAEFSTEELAEIDRYAKDADINLWAKSAERV"
 6465...6791
 /gene="AGR_C_2812"
 6465...6791
 /note="hypoetical protein"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_2812p"
 /protein_id="AAK87316.1"
 /db_xref="GI:15156613"
 /translation="MKKFTVILLATVLAAPMAQAQSRHDDRHRGVTVBERQVTTKKVI

VKKRWDRQRLSRERERNWVDRDYRYRHLAEPRDQWRVVDNQFLLINAVSLIV
 GLAAR"
 complement (6858...8288)
 /gene="AGR_C_2814"
 complement (6858...8288)
 /note="6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING"
 /codon_start=1
 /transl_table=11
 /product="AGR_C_2814p"
 /protein_id="AAK87317.1"
 /db_xref="GI:15156614"
 /translation="MEQAEIGLIGLVMSGLNALNIAEKNGKIAVFNTPVTRKFFA
 DAGELQGOIIPCEITEEFVAARPPRIIMIKAGDPVDQMEILKPHLANGDIMDA
 GNANRDFIIRPFDNLKDSGLTFPIGMVSGEGEAGRHGSPIMVGTEDSWKVERKVLTS
 ISAKTNDPFCVWLNGDAGHFVKTTHNGIEIADQMIAEIVLURDGLKMSAVESIA
 VPAENKRLMSYLIETEKVLRADPTIKRPMVDLILDKAGQKGTADIESALLAAKVG
 VAAETIAAARILSSQKDERAAEKIFGLPTLAAAPADKKAFTADIESALLAAKVG
 AYACQFATMSAASKEFNWLPMTPTIARWRAGCTIIRSEFDEITSFTKDPHVALIV
 TPAFSAIVKDTPDAPLRVVSVALSPVSALASALGYFDAYRRRGSGSANLQAORDF
 FGAHGFERTDGVDPHGWGSGADIF"
 BASE COUNT 1814 a 2570 c 2348 g 1740 t
 ORIGIN
 Query Match 6.9%; Score 164.4; DB 1; Length 8472;
 Best Local Similarity 57.2%; Pred. No. 2.7e-29;
 Matches 359; Conservative 0; Mismatches 236; Indels 33; Gaps 2;
 QY 3 ATACTCTTTTGAAGAAACCATGTACGATTGCTGACATTGTTGCTCTGGAAGGCTC 62
 DB 5669 ACACACCGCTTGAAGAAACCTCGGCGCGCTGGACCAGATCGTGCCTCGGCNAGCGC 5728
 QY 63 TTTACGTGGGTATTTCTTCTACGTCAGAGCTCAACGAGGAGGCGGTGAGTTTCATGG 122
 DB 5729 TCTATGTCGGCATCTCTCTCACTCAACTCGAAGCGCACCGCGGCGCGCTATCTCA 5788
 QY 123 CGGAGGAGGCGTCCCGCTTCTGATTTCATCAGCAGCTATTCATCATTAATCGTTGGG 182
 DB 5789 AGATCTCGGACGCTCTGATCATCCACGCGGAGCTATTTCGATGATCAACCGCTGGA 5848
 QY 183 TGGAGGAACCGGCGCATGACGCTGAGAACTTGTGCACTGAGTCCCAACAATGCTTGTG 242
 DB 5849 TCGAGGAACGCTCTTGTGCG-----ATACGCTGGAAGAACTCGGATCG 5893
 QY 243 CGCTCATCTCTTCTCACACATTCGCGAGGCGCTGCTACGAGCAAAATATCTCGATGGA 302
 DB 5894 GTCCATCGTCTTTTTCGCGCGCTGGCGGCGATGCTGACGACGAAATATCTGCGCGGTG 5953
 QY 303 TTCCAGAGGTTTCCCGCGCGCAGCGGTAAGTCCCTGCTCTGAGGCGCATGTTGAACGTGA 362
 DB 5954 TGGCGATGCGAGCGCTGCTTACAGAGCAAGTCACTCAACCCGCGCTTCTCTCAACGAGC 6013
 QY 363 ACAATATTGATGTGTCGCAAGCTCAATGACATCGCCGAGGAACGCGGCGAGTCACTTGTG 422
 DB 6014 GCAATGTGCAAAACATCCGCGCGCTGACAGCAATTCGCGAGCGCGCTGGCCAGACGCTGG 6073
 QY 423 CGCAGATGCGCTTGCATGGGTGCTGCGCGAGCAAGAGAGTACCGCGCGGATACCGTGA 482
 DB 6074 CGCAGATGCGCAATTCGCTGGGTCTCTGCGCGCGCGCG-----CATTA 6115
 QY 483 CCAGTGCATTGATTGCTGCTTCTGCTGAGTTCAGCTGAGCAGCTGGAACAACCTTGTTCCTCA 542
 DB 6116 CCTCAGCATTTGATTGGCGCAAGCGCTGTGCAACAGGTTCGAGGACTTCGCTGAAGCACTCG 6175
 QY 543 ACAACTTGGAGTTTCTGACGCGCGAGTTGGAGCGCATGATGAGATTTTCCCAACGACGCG 602
 DB 6176 ATATGCCAGTCTCTACCGAGGAGCTGGCGCAATTCACCGTTACGCCAAGGATCGCG 6235
 QY 603 GCATCAACATTTGGCGGAGGCGCACCGCA 630
 DB 6236 ATATCAACCTCTGGGCAAAATCTGCCGA 6263

Search completed: April 27, 2003, 00:07:17
Job time : 7473.94 secs

GenCore version 5.1.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:02:03 ; Search time 3163.33 Seconds
(without alignments)
12154.292 Million cell updates/sec

Title: US-09-105-117K-3
Perfect score: 2374
Sequence: 1 agataactctttggaagaaa.....gtacacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hct:*
9: gb_est1:*
10: gb_est2:*
11: gb_hct:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	84.6	3.6	657	14	BQ155035
C 2	73.2	3.1	3237	17	BH770958
C 3	72	3.0	860	17	AF075981
4	60.4	2.5	830	17	CNS01MQH
5	48.8	2.1	1101	17	CNS017SY
C 6	48.6	2.0	925	17	CNS0091P
					BQ155035 NF075E081
					BH770958 LLMGtag68
					AF075981 AF075981
					AL151258 Anopheles
					AL108460 Drosophil
					AL053013 Drosophil

7	44.4	1.9	644	14	BQ704105
C 8	44.4	1.9	664	17	CNS03K4J
9	42.8	1.8	532	10	AW927407
10	41.8	1.8	947	17	CNS077MV
11	41.8	1.8	1083	17	CNS079WN
12	41.6	1.8	558	14	BM712036
C 13	41.2	1.7	910	17	CNS006ON
14	41	1.7	617	14	BM729055
15	41	1.7	770	12	BE898495
16	41	1.7	861	12	BE740880
17	40.8	1.7	950	11	AY104298
C 18	40.6	1.7	436	17	AQ405681
C 19	40.2	1.7	613	10	BE498441
C 20	40	1.7	357	9	AI418751
C 21	40	1.7	431	9	AA452858
22	40	1.7	447	14	W91005
C 23	40	1.7	463	9	AA476500
C 24	40	1.7	504	9	AI159901
C 25	40	1.7	540	9	AI401605
C 26	40	1.7	564	9	AI820014
C 27	40	1.7	600	12	BG806166
C 28	40	1.7	916	17	CNS015QU
29	39.8	1.7	925	17	CNS0091P
C 30	39.6	1.7	1101	17	CNS00KK2
C 31	39.2	1.7	1667	11	AY108197
C 32	39.2	1.7	846	17	CNS010RJ
C 33	39.2	1.7	884	17	CNS006UO
C 34	38.8	1.6	557	10	AW787854
C 35	38.8	1.6	1101	17	CNS017SY
C 36	38.6	1.6	436	10	BE025218
C 37	38.6	1.6	513	14	BQ818660
C 38	38.6	1.6	828	17	CNS026XT
39	38.4	1.6	650	10	BB576857
40	38.4	1.6	839	17	CNS004NB
C 41	38	1.6	212	9	AA150978
C 42	38	1.6	486	9	AA149286
C 43	38	1.6	903	13	BG975548
44	37.8	1.6	550	17	BH389688
45	37.8	1.6	804	17	CNS01HIJ

ALIGNMENTS

RESULT 1
BQ155035/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ155035 657 bp mRNA linear EST 24-APR-2002
NF075E081R1F1067 Irradiated Medicago truncatula cDNA clone
NF075E081R 5', mRNA sequence.

BQ155035
BQ155035.1 GI:20292094

EST.
barrel medic.

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.

1 (bases 1 to 657)

Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

Medicago truncatula irradiated library

Unpublished (2001)

Contact: May GD

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7391

Fax: 580 221 7380

Email: gdmay@noble.org

Insert Length: 657 Std Error: 0.00

Plate: 075 row: E column: 08


```

QY 535 TTCACTCAACAACTTGGAGTTTCTGACGCCGAGTT 570
Db 1991 AGCTTTAGAACGCTGCGAATTACTGACGAAGAATT 1956

RESULT 3
LOCUS AF075981
DEFINITION AF075981 Salmonella typhimurium LT2, Lambda DASH II linear GSS 29-AUG-2000
ACCESSION AF075981
VERSION AF075981.1 GI:3320851
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
REFERENCE 1 (bases 1 to 860)
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
JOURNAL comparison to the Escherichia coli K12 genome
MEDLINE FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
COMMENT 92243757
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdsu.edu
Class: shotgun.
FEATURES
source
1..860
Location/Qualifiers
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="390-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 212 a 199 c 242 g 207 t
ORIGIN

Query Match 3.0%; Score 72; DB 17; Length 860;
Best Local Similarity 54.7%; Pred. No. 1.8e-09;
Matches 220; Conservative 0; Mismatches 160; Indels 22; Gaps 3;

QY 52 TCGAAAGGCTTTTACGTGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGAGCGGC 111
Db 18 TGGCAAGCGTTGTAGTGGGATCTTAACATCTCTGCGGATCTGCGCAGACAAGCTAT 77

QY 112 TCAGTTTCATGCGGAGGAGCGCTTCTTCTGATTCATCAGCCAAGCTATTTCATCAT 171
Db 78 CGATATTCATGAGGATCTCGGACGCTTGCCTGATTCATCAGCCTAATATTCGCTTTT 137

QY 172 TAATCGTTGGTGGAGAAACCGGGGATGACGGTGAGAACTTTGTCAGTCACTGCGCAA 231
Db 138 TGAGCGTTGGGTAGAGGACGGGC-----TGCTGGCGCTGTTCAGGA 179

QY 232 CAATGCTCTTGGCGTCATGCTTCTCACCATTGCGCAGGCGCTGCTCACGCAATAA 291
Db 180 AAAAAAGCTCGGTAGTATTGCTTCTCGCGCTGCGGCGGCGCA-GCTCACTGACCGTAA 238

QY 292 TCTCGATGAAATCCAGAGGGTTCCCGC---GCCAGCCAGGGTAAAGTCCCTGCTGAGG 348
Db 239 TTGGAATGTAATTCGGAAGATTCCCGCGCGGAGCGGAAGCCGTTCTTAACCCAGA 298

QY 349 CATGTTGAACGTGAACATATTGATGTGTCGCAAGCTCAATGACATCGCCAGGAACG 408
Db 299 ACAGATTACCGCGCAAACTGGAAGAAATTCGCGCGTTGAAATGAATGAATGCTGCGCG 358

QY 409 CGGGAGTCACATTGCGCAGATCGCGCTTCATGGGTGCTGGC 450
Db 359 GGGTCAAAATTTGATCAAAATGGCGCTCGCGCTGGATGCTGGC 400

CNS01MQH 830 bp DNA linear GSS 14-JUN-2001
Anopheles gambiae GSS T7 end of clone 22E24 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
AL151258
AL151258.1 GI:7011737
GSS.
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 830)
Genoscope.
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 830)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
Submitted (16-FEB-2000) BMMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
1..830
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="22E24"
/clone_lib="NotreDamel"
/note="end : T7"
BASE COUNT 207 a 231 c 227 g 156 t 9 others
ORIGIN

Query Match 2.5%; Score 60.4; DB 17; Length 830;
Best Local Similarity 50.7%; Pred. No. 4.2e-06;
Matches 142; Conservative 1; Mismatches 137; Indels 0; Gaps 0;

QY 1058 GCGGTTTCACCTTTTGTGTCATGGCGTCTTTCGTCGCATGACGGCAACCATACAGG 1117
Db 318 GCCAGTTTCAGGTTACTGCTCATCGCGCTTTTCAGCGCGCGGAATCCGTACGAGCAGA 377

QY 1118 TAAGCGATGCCACCCAGCGGATATATCGAGCAGATCGGCGCGCATTCGACAAAGA 1177
Db 378 AACGCCACGCGCGCCAGGTACCAACCCAGCAGCCAGGAGCTGTCATCAGCAGCGCG 437

QY 1178 TCAAGCCCAAGGTGCGCGGATGAACAAAGAGCGTCAGAAATTAACACACAGCAAGA 1237
Db 438 CTGCGGCCAAATCCCGCACAGATCAGACCAATCACTTACCGCATAGCAGGCA 497

QY 1238 ACCGCAATGATCTTCGCGCTTAATTCCTGTTTAAATCAACAGTACATTCGCGTCCG 1297
Db 498 ATCATCAGATGTCATGCGCGGATGCTGCTGATTCATCAGCAAGCATTTTCGCGCGC 557

QY 1298 ATGGACACTAAAGACTGCGCGCCCAAGAGCAGACTGTAA 1337
Db 558 AGGGAGAAATATAAATGACCTTAAGCAAGCCCTTGAM 597

RESULT 5
LOCUS CNS017SY
DEFINITION CNS017SY Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108460

```

```

VERSION      AL108460.1  GI:5628764
KEYWORDS     GSS.
SOURCE       Drosophila melanogaster.
ORGANISM     Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 1101)
AUTHORS      Direct Submission
TITLE        Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
              - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
              collaboration with the European Drosophila Genome Project (EDGP) -
              http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
              library (Dros BAC) was made by Alain Billaud at CSHP (Centre
              d'Etude du Polymorphisme Humain) with funding provided by a MRC
              project grant. The DNA was prepared from embryos by Alain Bucheton
              and Genevieve Payan. It has been constructed in the vector
              pBelOBAC11.
FEATURES     Location/Qualifiers
             1..1101
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="DrosBAC"
                /plasmid="pBelOBAC11"
                /note="end : SP6"
BASE COUNT   254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match      2.1%; Score 48.8; DB 17; Length 1101;
Best Local Similarity 14.5%; Pred. No. 0.012;
Matches 47; Conservative 156; Mismatches 121; Indels 0; Gaps 0;
QY 560 GAGCGGAGTTGGAGCGGATCGATGATTTCCACAGCGCGGCATCAATTTGGCG 619
Db 756 GAGGAGGGRGGGGCCAGSSVCGACASGASGSMCGVSSGSCSAGCCGCGVSC 815
QY 620 AAGCCACCATTCCTCAAAACCCCGGAAATCAACCATCAATCAGTTGATGGCCAAT 679
Db 816 SAVSASSSVMSKVASAVASVAVASGAVSSCRSSVSVASVSVSSVSS 875
QY 680 GCGGTATCAAACTCCAGCAGCAGCTGATCCAGCGGCACACCTTGGGCTGGACAGC 739
Db 876 SSSVWAAAVSASSSSASASSMAVAAAAVAVSSVSVASVSSSSSCSSSSASVVSAS 935
QY 740 GGGCGTGACAATCTGCTGCGCGGAACCCACCGCGGGAACAGATCAGGCTTGGCGG 799
Db 936 VAASASVSSSSSVSTSSASVSVASVMSAVVSSSASSSVSVSVVAVASAA 995
QY 800 AAGCGCCAGCGCGGAAATCCACCGTCCGCGTTCGCGGTATTGCGCGCGCGCGCG 859
Db 996 AAAAAAASSSASAVAVVSVSSSSSSSSSSSSSSSSSSSSSSSVSVSVSSSV 1055
QY 860 ATAAACAAACCGCTCAATAC 883 .
Db 1056 SVAVASASASVSVVAAAATAAM 1079

RESULT 6
CNS0091P/c    925 bp  DNA  linear  GSS 03-JUN-1999
LOCUS         Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION    BAC19D16 of RPCI-98 library from Drosophila melanogaster (fruit
              fly), genomic survey sequence.
ACCESSION     AL053013
VERSION       AL053013.1  GI:4934461
KEYWORDS      GSS.
SOURCE        Drosophila melanogaster.
ORGANISM      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

```

```

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 925)
AUTHORS      Direct Submission
TITLE        Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
              - Web : www.genoscope.cns.fr)
COMMENT      Determination of this BAC-end sequence was carried out as part of a
              collaboration with the Berkeley Drosophila Genome Project (BDGP).
              The BDGP is constructing a physical map of the Drosophila
              melanogaster genome using these BACs. For further information
              please see http://www.fruitfly.org The BDGP Drosophila
              melanogaster BAC library was prepared by Kazutoyo Osoegawa and
              Aaron Mamoser in Pieter de Jong's laboratory in the Department of
              Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
              NY. The library is named RPCI-98 and was constructed by partial
              EcoRI digestion of Drosophila DNA provided by the BDGP from the
              isogenic strain y2; cn bw sp, the same strain used for the BDGP's
              F1 and EST libraries. A more detailed description of the library
              and how to order individual BAC clones, the entire library, or
              filters for hybridization from the BACPAC Resource Center can be
              found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES     Location/Qualifiers
             1..925
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="BAC19D16"
                /clone_lib="RPCI-98"
                /note="Tend : TET3"
BASE COUNT   120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match      2.0%; Score 48.6; DB 17; Length 925;
Best Local Similarity 13.8%; Pred. No. 0.012;
Matches 49; Conservative 157; Mismatches 149; Indels 0; Gaps 0;
QY 681 CGGTATCAAACTCCAGCAGCAGCTGATCCAGCGGCACACCTTGGGCTGGACAGC 740
Db 925 CBSBCSCSCSBSBSSSSMTSSSBSCSSBSSSTSSBSSSSBSSSSSSSSSS 866
QY 741 GCGGTGACAATCTGCTGCGCGGAACCCACAGCGGGAACAGATCAGGCTTGGCGGA 800
Db 865 SSGTSACVKNASSCGCGCGGABCCWSSSSSCGASARGVKVRASGAGKRGGS 806
QY 801 AGCGCCAGCGCGGAAATCCACCGTCCGCGTTCGCCCGTATTGCGCGCGCGCGCG 860
Db 805 GGASASHSSSACBSBSSSSCSASCSWSASSSSSRSSRSGGAGSGASSSSSSSS 746
QY 861 TAAACAAACCGCTCCAAATACGATTCGGGTTCAACAGGTACAGCAGATTGCCATCA 920
Db 745 ASAGSVSVSSSSSSSSSSSVSCSVASMSCSBSSSSASASSSSSSSSSSSSSC 686
QY 921 ACATGGGCTTTACCAACACCGCTCTTATGACGCTCACCTCCACCGCGCGCGGTTGC 980
Db 685 TSWSCSTASMSAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 626
QY 981 GCGTGTGATGGCCACCGCGGAACCGCCCAAGGCGGTGTCTATCGGGCACGTTGC 1035
Db 625 SMSGGSGSVSVASSGMSVSSGGRSSGSGGSGGSGGSGSGSGSGSGSGSGS 571

RESULT 7
BQ704105
LOCUS         BQ704105
DEFINITION    946110C04.y1 946 - tassal primordium prepared by Schmidt lab Zea
              mays cDNA, mRNA sequence.
ACCESSION     BQ704105
VERSION       BQ704105.1  GI:21843524
KEYWORDS      EST.
SOURCE        Zea mays.
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 644)
Walbot.V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 946110 row: C column: 04.

FEATURES

source

1. 644
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassels primordium prepared by Schmidt
lab"
/tissue_type="tassels"
/dev_stages="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/note="Organ: tassels; Vector: HybriZAP; Site 1: ECGRI;
Site 2: Xhot; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanford prepared the cDNA
library in HybriZAP. Sample insert size range was 350 bp
to 3 kb with a 1 kb average."
118 a 239 c 180 g 107 t

BASE COUNT
ORIGIN

Query Match 1.9%; Score 44.4; DB 14; Length 644;
Best Local Similarity 46.0%; Pred. No. 0.18;
Matches 150; Conservative 0; Mismatches 176; Indels 0; Gaps 0;
QY 686 ATCAACATGCGCCAGCAGCGTGTATCCAGCGCACACCTTGGGGTGGACAGCGGGGT 745
Db 39 ATCCCAATATCGCAACGCGATGTCTCGCTGCGCCGATGTCTCTCGCGTGGCC 98
QY 746 GACATGTCTGCGCGGAAACCCACAGCGGGAACAGATCAGGCTTCCCGCAACGCG 805
Db 99 GCGCGCGCGCGAGCGGTCTACGCGGTCTAGCCAGCTCAGCGCGCGCGAAGCAG 158
QY 806 CCAGCGGGAATCCACGCTCGGTGTGCGGTATGCGCGCGCGCGCGGATTAAC 865
Db 159 CTACTCCCAAGTACGCGCTCCCGAAGGGGCTCATCCCGGACTCCGCTCTTACAGC 218
QY 866 ACAACGCGCTCCAAATACGATTCGGGTTCAACAGGTCTCAGCAGGATTCGCAATCAACATG 925
Db 219 TTCGAGAGGCGCACGGGCGCTTCGATTCACCTCTCGCGGCACTGTACGTCCACTTC 278
QY 926 GCGTTTACCAACCGGTGTCTATCGAGCTCAGCTCAGCTCCACCGCGACCGGTTGCGCGTG 985
Db 279 GCGTCCCACTCGTCTACTACGAGAGGACCAATACCGGAAGCTCTCCAAAGGCGCCATC 338
QY 986 TCAGTGGCCACCGCGAAGCGCCCA 1011
Db 339 TCGGACCTCTCCGCGTCCAGGCCAA 364

RESULT 8

CNS03K4J/C

LOCUS

DEFINITION CNS03K4J 664 bp DNA linear GSS 17-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
032B06 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION AL247708

VERSION AL247708.1 GI:7968720

KEYWORDS GSS; genome survey sequence.

SOURCE Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS

1 (bases 1 to 664)

Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.

Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE

2 (bases 1 to 664)

Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,

Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.

Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

JOURNAL

REFERENCE

3 (bases 1 to 664)

Genoscope.

Direct Submission

Submitted (12-APR-2000)

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetraodon nigroviridis

genome. For more information, please take a look at

http://www.genoscope.cns.fr/tetraodon.

FEATURES

source

1. 664

/organism="Tetraodon nigroviridis"

/db_xref="taxon:99883"

/clone="032B06"

/clone_lib="G"

/note="Genoscope sequence ID : COBG032DA03LP1-end : T7"

118 a 200 c 222 g 116 t 8 others

Query Match 1.9%; Score 44.4; DB 17; Length 664;

Best Local Similarity 48.2%; Pred. No. 0.18; 129; Indels 0; Gaps 0;

Matches 120; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 526 CAGCCTTGATTCACCTCAACAACTTGGAGTTTTCTGACCGCGAGTTGGAGCGCATCGATGA 585

Db 526 CGCGCGGGAAGACCCACACGCGGGGAGCTCCACGCGCGGGAGGYTCCACCGTGGG 467

QY 586 GATTTCCACAGCGCGCATCAACATTTGGGCGAAGCCACCGATTCCAAAACCCGCGA 645

Db 466 GGGAGCCTTCCGCGCGCGCGAGGCTCCATCGTGGAGAGCCACCGGCTCCGCGCGG 407

QY 646 AAACCTAACCCCATCAACATCAGTTTGTATGGCCCAATGGGTCTATCACTGCGCAGCAC 705

Db 406 ATCGTTGCGCGACCAACCAACAGCGCGCGGCTGTCCCGGCGCGCGCGCGCAAA 347

QY 706 GTTGATCCAGCGCGCACACCTTTGGGCGCTGGACAGCGGGCGGTGACAATGTCTGCGCGCGAA 765

Db 346 AGANCCGGGTGCTGCTCGGCTCCGTCGCGGGGCGCGCGCGCGCGCGCGCGCGA 287

QY 766 ACCCACCAG 774

Db 286 CGGACCTG 278

RESULT 9

AW927407

LOCUS

DEFINITION

AW927407

ACCESSION

AW927407

VERSION

AW927407.1

KEYWORDS

EST.

SOURCE

ORGANISM

Zea mays

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AW927407 532 bp mRNA linear EST 30-MAY-2000
945001E08.Y1 945 - Mixed adult tissues from Walbot lab, same as 707
(SK) Zea mays cDNA, mRNA sequence.

ACCESSION AW927407

VERSION AW927407.1

KEYWORDS EST.

SOURCE Zea mays

ORGANISM Zea mays

ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 532)

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 945001 row: E column: 08.

Location/Qualifiers

1. .532

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="945 - Mixed adult tissues from Walbot lab,

same as 707 (SK)"

/issue type="tassel, kernal, silk, husk, root, leaf"

/dev stage="fully-grown"

/lab_host="DH108"

/note="Organ: tassel, kernal, silk, husk, root, leaf;

Vector: pGAD10; Site 1: EcoRI; cDNA library from fully

differentiated maize tissues from an active Mutator plant.

Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,

root, leaf). Unidirectionally cloned. New library number

given to library 707 for additional sequencing."

BASE COUNT 85 a 206 c 152 g 89 t

ORIGIN

Query Match 1.8%; Score 42.8; DB 10; Length 532;

Best Local Similarity 45.7%; Pred. No. 0.5;

Matches 149; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 686 ATCAAACTGCCAGCAGCGTGTATCCAGCGGCACACCTTGGGGTGGACAGCGGGCGT 745

DB 7 ACCCCAAATATCCGAAGCGCAATGTCTCCGCTGCCCGCATGTCTCTCGCGCGTGGCG 66

QY 746 GACAATGCTCTCGCGGGAACCCACAGCGGGAACAGATCAGGCTTCCGCGAAGCGG 805

DB 67 GCCGCGCCCGCAGCGGTGTCAGCGGTGACGCCACGTCTCAGCGCGCGCGAAGCAC 126

QY 806 CCAGCGCGGAAATCCACCGCTCCCGGTGTCTCGCGTATTTGCGCGCGGACGCGCGATAAAC 865

DB 127 CTACTCCCAAGTACGGCTTCCGAGGGGCTCATCCGACTCCGTCGGCTCTACAGC 186

QY 866 ACAAACGGTCCAAATACGATTCGGGTTCAACAGGTCAAGCAGATTCGCAATCAACATG 925

DB 187 TTCAGCAGGCGCACGGCGGCTTCAGATCCACTCGCGCGCACCTGTAGTCCACTTC 246

QY 926 GGCTTTACCAACCCGCTCTTATCGACGCTCACTCCACCGCACCGGTTGCGCGTG 985

DB 247 GGCTCCACCTCTGTACTACGAGAGGACCAATACCGGCAAGCTCTCCAGGCGGCGATC 306

QY 986 TCAGTGGCCACCGCGCAACCGCCCAA 1011

DB 307 TCGGACTCTCCGGGTCCAGGCCAA 332

RESULT 10

CNS077MW

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CNS077MW 947 bp DNA linear GSS 08-JUL-2001

T3 end of clone BB0AA013All of library BB0AA from strain CBS 4732

of Pichia angusta, genomic survey sequence.

AL432893

AL432893.1 GI:12216307

GSS.

Pichia angusta.

Pichia angusta

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

1 (bases 1 to 947)

Souciot, J.L., Aigle, M., Artiguenave, F., Blandin, G.,

Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,

de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B.,

Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Poirier, S.,

Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,

Wincker, P., and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of

yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

20584711

11152876

2 (bases 1 to 947)

Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F.

and Dujon, B.

Genomic exploration of the hemiascomycetous yeasts: 13. Pichia

angusta

FEBS Lett. 487 (1), 76-81 (2000)

20584723

11152888

3 (bases 1 to 947)

Genoscope.

Direct Submission

Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,

2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :

secre@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen

yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces

exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,

Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia

angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,

Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to

5 kb were prepared and both extremities were sequenced. See

keywords for description of this sequence and for the sequence of

the other extremity of this insert.

FEATURES

Location/Qualifiers

1. .947

/organism="Pichia angusta"

/strain="CBS 4732"

/db_xref="taxon:4905"

/clone_lib="BB0AA013All"

/note="end : T3"

/note="similar to Saccharomyces cerevisiae ORF YPL088w [

similarity to aryl-alcohol dehydrogenases]"

/evidence=not experimental

BASE COUNT 301 a 181 c 206 g 259 t

ORIGIN

Query Match 1.8%; Score 41.8; DB 17; Length 947;

Best Local Similarity 51.3%; Pred. No. 1.2;

Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 2 GATATCTCTTTTGGAGAAACCATGTACGATTGCGTGACATTGTTGCGTCTCGAAGGCT 61

DB 338 GAGACACCATTTGAAGAAACGATGAGACATTTGATGATTTGTAAAGTCTCGAAGACT 397

QY 62 CTTTACGTGGGTATTTCTTCTTACGTCGAGTCACAGCGGAGCGGCTGAGTTTCATG 121

DB 398 CGTTATATTGGGGCTCGACATGAGAGCTTATCAATTTATCGAAATGCAACATGTTGCT 457

QY 122 CGGGAGGAGGCTGCCGCTTCTGATTTCATCAGCAAGCTATTCATCATTAATCGTTGG 181

DB 458 GAAAGACATGTTGTTGACAAAGTTCAATTTCCATGCAAAAGCTATTATTCCCTGTTCTTCCA 517

QY 182 GTGGAGGAA 190

DB 518 GAGGAAGAA 526

```

RESULT 11
CNS079WN      1083 bp      DNA      linear      GSS 08-JUL-2001
LOCUS          T3 end of clone XBB0AA001G01 of library XBB0AA from strain CBS 4732
DEFINITION    of Pichia angusta, genomic survey sequence.
ACCESSION     AL435837
VERSION       AL435837.1 GI:12219250
KEYWORDS      GSS.
SOURCE        Pichia angusta.
ORGANISM      Pichia angusta
              Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE     1 (bases 1 to 1083)
AUTHORS       Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
              Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
              de-Montigny, J., Dujon, B., Durand, P., Lepingle, A., Liorente, B.,
              Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Poirier, S.,
              Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
              Wincker, P., and Weissenbach, J.
              Genomic exploration of the hemiascomycetous yeasts: 1. A set of
              yeast species for molecular evolution studies
              FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE       20584711
PUBMED        11152876
REFERENCE     2 (bases 1 to 1083)
AUTHORS       Blandin, G., Liorente, B., Malpertuy, A., Wincker, P., Artiguenave, P.,
              and Dujon, B.
              Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
              angusta
              FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE       20584723
PUBMED        11152888
REFERENCE     3 (bases 1 to 1083)
AUTHORS       Genoscope.
              Direct Submission
              Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
              2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
              seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
              This GSS is part of a random genomic sequencing program of thirteen
              yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
              exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
              Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
              lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
              angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
              Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
              5 kb were prepared and both extremities were sequenced. See
              keywords for description of this sequence and for the sequence of
              the other extremity of this insert.
              Location/Qualifiers
                1..1083
                /organism="Pichia angusta"
                /strain="CBS 4732"
                /db_xref="taxon:4905"
                /clone="XBB0AA001G01"
                /clone_lib="XBB0AA"
                /note="end : T3"
                <40..>717
              misc_feature
                /note="similar to Saccharomyces cerevisiae ORF YPL088w [
                similarity to aryl-alcohol dehydrogenases ]"
                /evidence=not_experimental
BASE COUNT    326 a 214 c 234 g 308 t
ORIGIN
Query Match      1.8%; Score 41.8; DB 17; Length 1083;
Best Local Similarity 51.3%; Pred. NO. 1.3;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
Qy 2 GATACCTCTTTCGAGAACACCATGACGATTCGGTACATTTGGCTCTCGAAGGCT 61
      |||
Db 163 GAGACACCATGATGAGAACACGATGAGACGATTCGATGATTTGTGAAGTCTGGAAGACT 222
      |||
Qy 62 CTTTACGTGGGTATTCTTCTTCACGGTCCAGAGCTCACACGGCGGCGGCTGAGTTCATG 121

```

```

Db 223 CGTTATTTGGGGCTCGACATGAGAGCTTATCAATTTATCGAATGCAACATGTTGCT 282
      |||
Qy 122 GCGGAGGAGGGCTGCCGGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATCGTTGG 181
      |||
Db 283 GAAAAGCATGGTTGGACAAAGTTTCATTTCCATGCAAAAGCTATTATTTCCCTGTTCTTCAGA 342
      |||
Qy 182 GTGGAGGAA 190
      |||
Db 343 GAGGAAGAA 351
      |||

RESULT 12
LOCUS          BM712036
DEFINITION    UI-E-DWI-ahc-g-16-0-UI-r1 UI-E-DWI Homo sapiens cDNA clone
              UI-E-DWI-ahc-g-16-0-UI 5', mRNA sequence.
ACCESSION     BM712036
VERSION       BM712036.1 GI:19025294
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 558)
AUTHORS       Bonaldo, M.F., Lennon, G., and Soares, M.B.
              Normalization and subtraction: two approaches to facilitate gene
              discovery
              Genome Res. 6 (9), 791-806 (1996)
JOURNAL       97044477
MEDLINE
COMMENT       Contact: Soares, MB
              Program for Rat Gene Discovery and Mapping
              University of Iowa
              451 Eckstein Medical Research Building Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9565
              Email: msoares@blue.weeg.uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com)
              The following repetitive elements were found in this cDNA
              sequence: 143-166, >GC-rich#low_complexity
              Seq primer: M13 Reverse.
              Location/Qualifiers
                1..558
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="UI-E-DWI-ahc-g-16-0-UI"
                /clone_lib="UI-E-DWI"
                /tissue_type="lens"
                /dev_stage="adult"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /note="Organ: eye; Vector: pTT3-Pac (Pharmacia) with a
                modified polylinker; Site: 1: EcoR I; Site 2: Not I;
                UI-E-DWI is a normalized cDNA library containing the
                following tissue(s): lens. The library was constructed
                according to Bonaldo, Lennon and Soares, Genome Research,
                6:791-806, 1996. First strand cDNA synthesis was primed
                with an oligo-dT primer containing a Not I site. Double
                stranded cDNA was ligated to an EcoR I adaptor, digested
                with Not I, and cloned directionally into pTT3-Pac
                vector. The oligonucleotide used to prime the synthesis of
                first-strand cDNA contains a library tag sequence that is
                located between the Not I site and the (dT)18 tail. The
                sequence tag for this library is CGATTACCGA. This library
                was created for the program, Gene Discovery in the Visual
                System, supported by National Eye Institute (NEI)."
BASE COUNT    113 a 169 c 185 g 90 t
ORIGIN

```

[illegible]

RESULT 14
 BM729055
 LOCUS
 DEFINITION
 UI-E-EO1-aiw-a-06-0-UI.r1 UI-E-EO1 Homo sapiens CDNA clone
 UI-E-EO1-aiw-a-06-0-UI 5', mRNA sequence.
 ACCESSION
 BM729055
 VERSION
 BM729055.1 GI:19050388
 KEYWORDS
 EST. .
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 617)
 Nardito,M.F., Lennon,G. and Soares,M.B.
 AUTHORS
 Normalization and subtraction: two approaches to facilitate gene
 TITLE
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 MEDLINE
 97044477
 CONTACT
 Contact: Soares, MB

Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 148-171, >GC rich#Low_complexity
Seq primer: M13 Reverse.

```

FEATURES
source
seq primer: W13 Reverse.
Location/Qualifiers
1. .617
/organism="Homo sapiens"
/db xref="taxon:9606"
/clone="UI-E-EO1-aiw-a-06-0-UI"
/clone_lib="UI-E-EO1"
/tissue_type="fetal eye"
/dev stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EO1 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was

```


constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-fac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

```

BASE COUNT      124 a  189 c  211 g  93 t
ORIGIN

Query Match      1.7%; Score 41; DB 14; Length 617;
Best Local Similarity 46.9%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 585 AGATTTCCACGACGCGGCATCAACATTTGGGCGAAGCCACCGATTCAAAACCCGCG 644
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 265 ACATCGCAACGAGGACTGCATCAAGGTGGCGCGCTGCTCAACACGCCCTTCTACTGG 324
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 645 AAACTAACCCATCAACATCAGTTTGATGGCCAAATGGGTGTCATCAACATGCGCACGACGA 704
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 325 AGAACCTGCACCTTCCACATCGAGGCGAAGGACACGCACTACTTCAACAGACCAACACGC 384
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 705 CGTTGATCCAGCGCCACACCTTGGGCTGGACAGCGGGCTGCAATGCTGCTGCGCGGA 764
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 385 CCAGAGCGACTTGGGACACGCTGCGTTGACAGCGGCCCAAGGCGCTGGAGAACGGCA 444
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 765 AACCCACGACGCGGAACAGATCAGGTTGCGCGCAACGCGCGCGGCAAAATCCACC 824
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 445 TCAACGTGACGGTGTGCGACGTCCACACGCTGTGTAAGCGCAGGACGCGAGTTGCGCG 504
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 825 GTCCGCTGTCGCGTATTTGGCGCGCCGACGCGC 857
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 505 ACCTGGAGATGCAGTTTGGCGCGCTGCGCGTGC 537
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

RESULT 15
BE898495
LOCUS      601681396F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3951622 5',
          mRNA sequence.
ACCESSION BE898495
VERSION    BE898495.1 GI:10365029
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 770)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTB/DTF
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCW819 row: p column: 23
            High quality sequence stop: 765.
            Location/Qualifiers
            1..770
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3951622"
            /clone_lib="NIH_MGC_9"
            /tissue_type="adenocarcinoma cell line"

```

FEATURES
source

```

/lab host="DH10B (phage-resistant)"
/note="organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      147 a  245 c  277 g  101 t
ORIGIN

```

```

Query Match      1.7%; Score 41; DB 12; Length 770;
Best Local Similarity 46.9%; Pred. No. 1.9;
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 585 AGATTTCCACGACGCGGCATCAACATTTGGGCGAAGCCACCGATTCAAAACCCGCG 644
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 113 ACATCGCCAAACGAGGACTGCATCAAGGTGGCGCGCTGCTCAACAAAGCCTTCTACTGG 172
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 645 AAACTAACCCATCAACATCAGTTTGATGGCCAAATGGGTGTCATCAACATGCGCACGACGA 704
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 173 AGAACCTGCACCTTCCACATCGAGGCGAAGGACACGCACTACTTCAACAGACCAACACGC 232
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 705 CGTTGATCCAGCGCCACACCTTGGGCTGGACAGCGGGCTGCAATGCTGCTGCGCGGA 764
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 233 CCGAGAGCGACTTGGGCGACGCTGCGTTGACACGCGCGCGCTGGAGAACGGCA 292
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 765 AACCCACGACGCGGAACAGATCAGGTTGCGCGCAACGCGCGCGGCAAAATCCACC 824
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 293 TCAACGTGACGGTGTGCGACGTCCACACGCTGTGTAAGCGCAGGACGCGAGTTGCGCG 352
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 825 GTCCGCTGTCGCGTATTTGGCGCGCCGACGCGC 857
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 353 ACCTGGAGATGCAGTTTGGCGCGCTGCGCGTGC 385
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: April 27, 2003, 00:51:42
Job time : 3177.33 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 17:06:23 ; Search time 522.424 Seconds
(without alignments)
10233.535 Million cell updates/sec

Title: US-09-105-117K-3
Perfect score: 2374
Sequence: 1 agatactctttggaagaaa.....gtaacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2374	100.0	2374	18 AAT96816	DNA encoding LysG,
2	2374	100.0	349980	22 AAF68528	C glutamicum codin
3	993	41.8	993	22 AAF71777	Corynebacterium gl
4	993	41.8	993	23 AAS96096	C. glutamicum gene
5	870	36.6	870	22 AAF68421	C glutamicum codin
6	822	34.6	822	22 AAF71779	Corynebacterium gl
7	822	34.6	822	23 AAS96098	C. glutamicum gene
8	708	29.8	708	22 AAF68420-	C glutamicum codin
9	698.6	29.4	1568	22 AAF45375	C. thermoaminogene

10	627	26.4	627	22	AAH68419	C glutamicum codin
11	485.2	20.4	993	22	AAF68077	Corynebacterium gl
12	121.2	5.1	15239	17	AAT33536	BCG deletion regio
C 13	121.2	5.1	4403765	22	AAI99683	Mycobacterium tube
C 14	121.2	5.1	4411529	22	AAI99682	Mycobacterium tube
C 15	101	4.3	1200	22	AAF71729	Corynebacterium gl
C 16	100.8	4.2	5541	15	AAQ55755	Escherichia coli g
17	93.2	3.9	988	24	ABK72786	Bacillus lichenifo
18	82.6	3.5	894	23	AAS52558	E. coli DNA for ce
19	79.4	3.3	23128	23	AAS59552	Propionibacterium
C 20	78	3.3	1077	22	AAH66360	C glutamicum codin
21	76.2	3.2	894	23	AAS56007	Salmonella typhi D
22	73	3.1	903	23	AAS51548	Pseudomonas aerugi
23	68.8	2.9	2365589	24	ABA90521	Genomic sequence o
24	64.4	2.7	1452	23	AAS88994	DNA encoding novel
25	64	2.7	1277	24	AAH31869	Lactobacillus tham
C 26	60.4	2.5	636	21	AAAS2691	Escherichia coli y
27	57.6	2.4	762	23	AAS54006	Klebsiella pneumon
C 28	47.4	2.0	7521	23	AAS71378	DNA encoding novel
C 29	47.4	2.0	7521	23	AAS94251	DNA encoding novel
C 30	45.4	1.9	45613	22	AAF28535	Genomic fragment #
C 31	45.2	1.9	4915	20	AAH84317	Human breast cell
32	43.8	1.8	1973	22	ABA44805	Stealth virus nucl
33	43.8	1.8	1973	22	ABA55261	Human fetal liver
34	43.8	1.8	1973	22	ABA25005	Probe #3471 for ge
35	43.8	1.8	1973	22	AAK03514	Human brain expres
36	43.8	1.8	1973	22	AAK28970	Human bone marrow
37	43.8	1.8	1973	22	AAI13556	Probe #3489 for ge
38	43.8	1.8	1973	22	AAI34918	Probe #3604 used t
39	43.8	1.8	1973	22	AAI03446	Probe #3437 used t
40	43.8	1.8	1973	24	ABS03504	Human genome-deriv
C 41	41.6	1.8	2842	21	ABD00334	Rice raffinose syn
42	41	1.7	2157	22	AAH14096	Human cDNA sequenc
43	41	1.7	3270	22	AAH14671	Human cDNA sequenc
44	41	1.7	3614	22	AAH14183	Human cDNA sequenc
C 45	40.4	1.7	1075	22	ABA48937	Human breast cell

ALIGNMENTS

RESULT 1
AAT96816/c
ID AAT96816 standard; DNA; 2374 BP.
XX
AC AAT96816;
XX
DT 12-MAR-1998 (first entry)
DE DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.
XX
KW LysG, LysE; ORF3; lysine transport; regulatory protein; export protein;
KW Microbial production; amino acid; animal feed additive; ds.
XX
OS Corynebacterium glutamicum.
XX
FH Key Location/Qualifiers
FT CDS complement (82..954)
FT FT /*tag= a
FT FT /label= LysG
FT FT 1016..1726
FT FT /*tag= b
FT FT /label= LysE
FT FT complement (1723..2373)
FT FT /*tag= c
FT FT /label= orf3

DE19548222-A1.

26-JUN-1997.

22-DEC-1995; 95DE-1048222.

PR 22-DEC-1995; 95DE-1048222.
 XX (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 XX Eggeling L, Sahm H, Vrljic M;
 XX
 XX WPI; 1997-333867/31.
 DR P-PSDB; AAW37714-16.
 DR
 XX
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT Corynebacterium
 XX
 PS Claim 23 and 26; Page -; 16pp; German.
 XX
 CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine,
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.
 CC NB. This sequence has been created from the information given in table 2
 CC of the specification.
 XX
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;
 Query Match 100.0%; Score 2374; DB 18; Length 2374;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGATACCTCTTGGAGAAACCATGTACGATTCGCGACATGTTGCGCTCGGAAGGC 60
 DB 2374 AGATACCTCTTGGAGAAACCATGTACGATTCGCGACATGTTGCGCTCGGAAGGC 2315
 QY 61 TCTTTACGTGGGTATTTCTTCTACCGTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 120
 DB 2314 TCTTTACGTGGGTATTTCTTCTACCGTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 2255
 QY 121 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCCAGCTATTCATTAATCGTTG 180
 DB 2254 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCCAGCTATTCATTAATCGTTG 2195
 QY 181 GGTGAGGAAACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTGCCAACATGTCT 240
 DB 2194 GGTGAGGAAACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTGCCAACATGTCT 2135
 QY 241 TGGCGTCATTGCTTTCTCACCACCTTGGCGAGGGCTCTGCTCACGGACAAATATCTCGATGG 300
 DB 2134 TGGCGTCATTGCTTTCTCACCACCTTGGCGAGGGCTCTGCTCACGGACAAATATCTCGATGG 2075
 QY 301 AATTCCAGAGGGTTCCCGCGCCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 360
 DB 2074 AATTCCAGAGGGTTCCCGCGCCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 2015
 QY 361 GAACAATATTGATGTGTCGACAGCTCAATGATCATCGCCCGAAGACCGGGCAGTCACT 420
 DB 2014 GAACAATATTGATGTGTCGACAGCTCAATGATCATCGCCCGAAGACCGGGCAGTCACT 1955
 QY 421 TGGCGAGATGGCGCTTGATGGGTGCTGCGCGAGCAAGAGAGTACGCGCGGATACCGT 480
 DB 1954 TGGCGAGATGGCGCTTGATGGGTGCTGCGCGAGCAAGAGAGTACGCGCGGATACCGT 1895
 QY 481 GACCAAGTCATTGATGGTGTCTGCTGATGAGCAGCTGGAACAGCCTTGATCACT 540
 DB 1894 GACCAAGTCATTGATGGTGTCTGCTGATGAGCAGCTGGAACAGCCTTGATCACT 1835
 QY 541 CAACAACCTTGAGTGTCTGACGCGGAGTTGGAGGCGATCGATCAGATTTTCCACGACGC 600

DB 1834 CAACAACCTTGAGTGTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACGACGC 1775
 QY 601 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAAACCCCGGAAAATAAATCCCATCAA 660
 DB 1774 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAAACCCCGGAAAATAAATCCCATCAA 1715
 QY 661 CATCAGTTTGTATGGCCAAATCGGTGATCACTCACTGCGCAGCAGAGCTTCATCCAGGCGCA 720
 DB 1714 CATCAGTTTGTATGGCCAAATCGGTGATCACTCACTGCGCAGCAGAGCTTCATCCAGGCGCA 1655
 QY 721 CACCTTTGGGGCTGGACAGCGGGCGGTGACAAATCTGCTGCGCGGAAAACCCACAGCGGAAA 780
 DB 1654 CACCTTTGGGGCTGGACAGCGGGCGGTGACAAATCTGCTGCGCGGAAAACCCACAGCGGAAA 1595
 QY 781 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAAATCCACCGTCCGGTGTCCCGTA 840
 DB 1594 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAAATCCACCGTCCGGTGTCCCGTA 1535
 QY 841 TTGGCGCGCGAGCGCGCGATTAACACAAACCGCTCCAAATACGCAATTCGGGTTCACCA 900
 DB 1534 TTGGCGCGCGAGCGCGCGCGATTAACACAAACCGCTCCAAATACGCAATTCGGGTTCACCA 1475
 QY 901 GGTTCAGCAGATTCGCTCAACATGCGGCTTTACCCAAACCCGCTGCTTATCCAGCTCAC 960
 DB 1474 GGTTCAGCAGATTCGCTCAACATGCGGCTTTACCCAAACCCGCTGCTTATCCAGCTCAC 1415
 QY 961 CTCACCCCGACCCCGGTTCGCGGTGTGAGTGGCCACCGCGGAAACCCCAAGGGGTGTC 1020
 DB 1414 CTCACCCCGACCCCGGTTCGCGGTGTGAGTGGCCACCGCGGAAACCCCAAGGGGTGTC 1355
 QY 1021 ATCGGCGCAGGTGTGTTCTGTTTCTCATGATCTGTGGCGCTTCCACCTTGTGTGAT 1080
 DB 1354 ATCGGCGCAGGTGTGTTCTGTTTCTCATGATCTGTGGCGCTTCCACCTTGTGTGAT 1295
 QY 1081 GCGGTCTTTTCGCTGCCATGACGCGGAAACCATTAACAGGTAAAGCGATGCCACCCAGCGCAT 1140
 DB 1294 GCGGTCTTTTCGCTGCCATGACGCGGAAACCATTAACAGGTAAAGCGATGCCACCCAGCGCAT 1235
 QY 1141 AATATCGACACGATCGCGCGGCTTGGACAAAAGATCAACGCCCAAGGTGCGCGCAT 1200
 DB 1234 AATATCGACACGATCGCGCGGCTTGGACAAAAGATCAACGCCCAAGGTGCGCGCAT 1175
 QY 1201 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAGAACCGCAATGAGTCTTTCGGCTT 1260
 DB 1174 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAGAACCGCAATGAGTCTTTCGGCTT 1115
 QY 1261 AATTCCTTGTAAATCACCAGTACATTTCTGCGGTCCGATGAGTAAAGAGTGGCCGC 1320
 DB 1114 AATTCCTTGTAAATCACCAGTACATTTCTGCGGTCCGATGAGTAAAGAGTGGCCGC 1055
 QY 1321 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
 DB 1054 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 995
 QY 1381 AGTAAATGATGGTCTTAAATGATGTTTAAATATAGTTCATGAACCCCAATCAATGGA 1440
 DB 994 AGTAAATGATGGTCTTAAATGATGTTTAAATATAGTTCATGAACCCCAATCAATGGA 935
 QY 1441 CACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAAAGCGGCTCTTAGCCCTTTCCAT 1500
 DB 934 CACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAAAGCGGCTCTTAGCCCTTTCCAT 875
 QY 1501 TTCCCTCTCGCGGTGAGTCAGCGGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
 DB 874 TTCCCTCTCGCGGTGAGTCAGCGGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 815
 QY 1561 GGTATCGCGCAACCAACCGGCGGAGAACCGAGGCGGTGAGTCTTGTGCAAGCAGC 1620
 DB 814 GGTATCGCGCAACCAACCGGCGGAGAACCGAGGCGGTGAGTCTTGTGCAAGCAGC 755
 QY 1621 GCGGAAAATGGTGTCTGCTGCAAGCAGAAACTAAAGCGCAACTATCTGGAGCGCTTGTCTGA 1680
 DB 754 GCGGAAAATGGTGTCTGCTGCAAGCAGAAACTAAAGCGCAACTATCTGGAGCGCTTGTCTGA 695

QY 1681 AATCCCGTTAAACCATGCCATCAACGAGATTCGCTATCAATGTTTCTCCCGTGT 1740
 |||||
 Db 694 AATCCCGTTAAACCATGCCATCAACGAGATTCGCTATCAATGTTTCTCCCGTGT 635
 |||||
 QY 1741 CAACGAGTAGCTTCTTGGGTGGAGAACGCTCAGCTGCTTGGAGATGAACGCA 1800
 |||||
 Db 634 CAACGAGTAGCTTCTTGGGTGGAGAACGCTCAGCTGCTTGGAGATGAACGCA 575
 |||||
 QY 1801 CACATATCTTCTCGCGGTGGAGATGTTTGGAGCGGTAAACCGTGAAGCTAATCC 1860
 |||||
 Db 574 CACATATCTTCTCGCGGTGGAGATGTTTGGAGCGGTAAACCGTGAAGCTAATCC 515
 |||||
 QY 1861 CBTGGCGGATGTAAGTAGTAGAATCTTGAACCATCGCGCACTTGGCCCATTCGAACCCG 1920
 |||||
 Db 514 CBTGGCGGATGTAAGTAGTAGAATCTTGAACCATCGCGCACTTGGCCCATTCGAACCCG 455
 |||||
 QY 1921 CTCATTTGGGATGCTTACATGTTTCATGGGAACCTAGATTGGCTGCGATGCCGCTTT 1980
 |||||
 Db 454 CTCATTTGGGATGCTTACATGTTTCATGGGAACCTAGATTGGCTGCGATGCCGCTTT 395
 |||||
 QY 1981 AGCTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGGAGCGCGCTGATGCTCC 2040
 |||||
 Db 394 AGCTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGGAGCGCGCTGATGCTCC 335
 |||||
 QY 2041 TGTGGGCGGAGCGGCTATCCATGTCCTCGTGGCGAAGTCTTGTGAGGCAATTCG 2100
 |||||
 Db 334 TGTGGGCGGAGCGGCTATCCATGTCCTCGTGGCGAAGTCTTGTGAGGCAATTCG 275
 |||||
 QY 2101 CCGAGCGCTTGTGGGACTTCTCCGAACCCAGCTGCTCCCATGCTAAAGCAGG 2160
 |||||
 Db 274 CCGAGCGCTTGTGGGACTTCTCCGAACCCAGCTGCTCCCATGCTAAAGCAGG 215
 |||||
 QY 2161 AGAAGTAGCTCTCTCGATGAGATACCATGTGACACACCGATGATTGGACGATGGG 2220
 |||||
 Db 214 AGAAGTAGCTCTCTCGATGAGATACCATGTGACACACCGATGATTGGACGATGGG 155
 |||||
 QY 2221 CTTGAATCTAGATCTCTAGCTAGACTCAGACGCGCTGTTGATGACGCAATCGAGG 2280
 |||||
 Db 154 CTTGAATCTAGATCTCTAGCTAGACTCAGACGCGCTGTTGATGACGCAATCGAGG 95
 |||||
 QY 2281 ATTGGGCTTGTACTTCTGAAAGGTTTCAAGGTTTCTTCTTCCCGCGAGGAA 2340
 |||||
 Db 94 ATTGGGCTTGTACTTCTGAAAGGTTTCAAGGTTTCTTCTTCCCGCGAGGAA 35
 |||||
 QY 2341 TTGGGCGGAGAGTAACACCTTCAGCAATGG 2374
 |||||
 Db 34 TTGGGCGGAGAGTAACACCTTCAGCAATGG 1

RESULT 2 AAH68528

ID AAH68528 standard; DNA; 349980 BP.

XX AC AAH68528;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 7063.

XX KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
 XX KW organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT
 PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from corynebacterium, and identifying a homologue of a gene derived
 CC from corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a nucleic acid described
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

SQ Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;

Query Match 100.0%; Score 2374; DB 22; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATACTCTTTGGAGAAACCATGTACGATTTGGTGACATTTGTTGCTTGGAAAGCG 60
 |||||

Db 127595 AGATACTCTTTGGAGAAACCATGTACGATTTGGTGACATTTGTTGCTTGGAAAGCG 127654
 |||||

QY 61 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCAGCGAGCGGCTGAGTTTCAAT 120
 |||||

Db 127655 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCAGCGAGCGGCTGAGTTTCAAT 127714
 |||||

QY 121 GCGGAGAGGCGCTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 180
 |||||

Db 127715 GCGGAGAGGCGCTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 127774
 |||||

QY 181 GGTGAGGAAACCGGCGGATGACGGTGAGAACTTGTTCAGTCAGTCCCAAAATGGTCT 240
 |||||

Db 127775 GGTGAGGAAACCGGCGGATGACGGTGAGAACTTGTTCAGTCAGTCCCAAAATGGTCT 127834
 |||||

QY 241 TGGCGTCAATGCTTTCTCACCATTGCGCAGGCGCTGTCTACGGACAAATATCTCGATGG 300
 |||||

Db 127835 TGGCGTCAATGCTTTCTCACCATTGCGCAGGCGCTGTCTACGGACAAATATCTCGATGG 127894
 |||||

QY 301 AATTCCAGAGGTTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACTG 360
 |||||

Db 127895 AATTCCAGAGGTTCCCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACTG 127954
 |||||

QY 361 GAAACAATATTGATATGGTCCCAAGCTCAATGACATGCCCGAGGAAACGCGGCGAGTCACT 420
 |||||

Db 127955 GAAACAATATTGATATGGTCCCAAGCTCAATGACATGCCCGAGGAAACGCGGCGAGTCACT 128014
 |||||

QY 421 TGGCAGATGCGCTTGCATGGGTGCTCGCGAGCAAGAGAGTACGCGCGGATACCGT 480
 |||||

Db 128015 TGGCAGATGCGCTTGCATGGGTGCTCGCGAGCAAGAGAGTACGCGCGGATACCGT 128074
 |||||

QY 481 GACCAGTGCATTTGATGGTGTCTGTCAGTTGACAGCTGACAAACGCTTGAATCACT 540
 |||||

Db 128075 GACCAGTGCATTTGATGGTGTCTGTCAGTTGACAGCTGACAAACGCTTGAATCACT 128134
 |||||

QY 541 CAACAACCTTGGAGTTTCTGAGCGGAGTTGGAGCGGATCGATGAGATTTCCCGAGGCG 600
 |||||

|||||
Db 128135 CAACAATGGAGTTTCTGACGCGAGTTGGAGCGATCGATGAGATTTCCACGACGC 128194
QY 601 CGGCATCAACATTTGGGCGAAGCCACCGATTCACAAAACCCGCGAAACCTAACCCATCAA 660
Db 128195 CGGCATCAACATTTGGGCGAAGCCACCGATTCACAAAACCCGCGAAACCTAACCCATCAA 128254
QY 661 CATCAGTTGATGCCCAATCGGTCATCAACAATGCGACACGATGATCAGCGCCA 720
Db 128255 CATCAGTTGATGCCCAATCGGTCATCAACAATGCGACACGATGATCAGCGCCA 128314
QY 721 CACTTTGGGCGTGCACAGCGGCGTGCACATGCTGTGCGCGAAACCCACAGCGGGA 780
Db 128315 CACTTTGGGCGTGCACAGCGGCGTGCACATGCTGTGCGCGAAACCCACAGCGGGA 128374
QY 781 CCAGATCAGGCTTTGCCGCGAACCGCGCAGCGGCGAAATCCACCGTCCGGTGTGCGCGTA 840
Db 128375 CCAGATCAGGCTTTGCCGCGAACCGCGCAGCGGCGAAATCCACCGTCCGGTGTGCGCGTA 128434
QY 841 TTGCGGCGCGACGCGCGGATTAACACAAACGCGTCCAAATACGATTCGGGTTCAACCA 900
Db 128435 TTGCGGCGCGACGCGCGGATTAACACAAACGCGTCCAAATACGATTCGGGTTCAACCA 128494
QY 901 GGTGACGACGATTTGCCATCAACATGGGCTTTACCCAAACCGGCTGCTTATCGACGCTCAC 960
Db 128495 GGTGACGACGATTTGCCATCAACATGGGCTTTACCCAAACCGGCTGCTTATCGACGCTCAC 128554
QY 961 CTCACCCGACCGGTTGCGCGTGTGAGTGGCCACCGCGAAACCCGCCAAAGCGGTGTC 1020
Db 128555 CTCACCCGACCGGTTGCGCGTGTGAGTGGCCACCGCGAAACCCGCCAAAGCGGTGTC 128614
QY 1021 ATCGGCGACGGTTGGTCTGTTCTTCTCAATGATCTGCGGCTTCCACCTTGTGTCAT 1080
Db 128615 ATCGGCGACGGTTGGTCTGTTCTTCAATGATCTGCGGCTTCCACCTTGTGTCAT 128674
QY 1081 GCGGCTCTTTGCGTGCCATGACGCGCAACACCAAGATGAGCGATGCCACCCAGCGCAT 1140
Db 128675 GCGGCTCTTTGCGTGCCATGACGCGCAACACCAAGATGAGCGATGCCACCCAGCGCAT 128734
QY 1141 AATATGAGCAGCATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTGTGCGCGAT 1200
Db 128735 AATATGAGCAGCATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTGTGCGCGAT 128794
QY 1201 GAACAAAGACGTCAGAAATTAACACACAGAGAACCGCAATGAGTCTTTCGCGCTT 1260
Db 128795 GAACAAAGACGTCAGAAATTAACACACAGAGAACCGCAATGAGTCTTTCGCGCTT 128854
QY 1261 AATTCTTGTGTTAATCACCAGTACATCTCGGTCGATGAGCAGTAAGAGCTGGCCCC 1320
Db 128855 AATTCTTGTGTTAATCACCAGTACATCTCGGTCGATGAGCAGTAAGAGCTGGCCCC 128914
QY 1321 CAAAAGCAGACCTGTAAATGAGATTTCCATGATCACCATCGTGAACCTATGGAAGTACTTA 1380
Db 128915 CAAAAGCAGACCTGTAAATGAGATTTCCATGATCACCATCGTGAACCTATGGAAGTACTTA 128974
QY 1381 AGTAAATGATGGTCTTAAATGATGGTTTAAATGATGGTTTAAATGATGGTTTAAATGATGG 1440
Db 128975 AGTAAATGATGGTCTTAAATGATGGTTTAAATGATGGTTTAAATGATGGTTTAAATGATGG 129034
QY 1441 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGGCTTCTTAGCCCTTTCCAT 1500
Db 129035 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGGCTTCTTAGCCCTTTCCAT 129094
QY 1501 TTCCCTTCGCGGTGATGACGCGGTTAAAGTCTTCGAGATCACTGCGGTGAGTGT 1560
Db 129095 TTCCCTTCGCGGTGATGACGCGGTTAAAGTCTTCGAGATCACTGCGGTGAGTGT 129154
QY 1561 GGTATCGCGCACCAACCGGCCAAGCAACCGAGCGGTTGAGTCTTCGCAAGCAGC 1620
Db 129155 GGTATCGCGCACCAACCGGCCAAGCAACCGAGCGGTTGAGTCTTCGCAAGCAGC 129214
QY 1621 GCGGAAATGGTGTGTCGCAAGCAGAACTTAAAGCGCAACTATCTGAGCGCTTGTCTGA 1680
|||||

Db 129215 GCGGAAATGGTGTGTCGCAAGCAGAAACTAAAGCGCAACTATCTGAGCGCTTGTCTGA 129274
QY 1681 AATCCGTTAACCATCGCCATCAACGAGATTCGTATCCACATGTTTCTTCCCGTGT 1740
Db 129275 AATCCGTTAACCATCGCCATCAACGAGATTCGTATCCACATGTTTCTTCCCGTGT 129334
QY 1741 CAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTGCGCTTGGAGATGAAGCGCA 1800
Db 129335 CAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTGCGCTTGGAGATGAAGCGCA 129394
QY 1801 CACATATCTTCTGCTGCGCGTGGAGATGTTTGGAGCGGTAAACCCGTGAAGCTAATCC 1860
Db 129395 CACATATCTTCTGCTGCGCGTGGAGATGTTTGGAGCGGTAAACCCGTGAAGCTAATCC 129454
QY 1861 CTTGCGGATGTGAAGTAGTAGAACTTTGGAAACCATGCGCCACTTTGGCCATTTGCAACCCC 1920
Db 129455 CTTGCGGATGTGAAGTAGTAGAACTTTGGAAACCATGCGCCACTTTGGCCATTTGCAACCCC 129514
QY 1921 CTCATTTGGGATGCTTACATGTTGATGGGAACTAGATTTGGGCTGCGATGCCGTCTT 1980
Db 129515 CTCATTTGGGATGCTTACATGTTGATGGGAACTAGATTTGGGCTGCGATGCCGTCTT 129574
QY 1981 ACGCTTTCGCTCCAAAGATGTCTTCAAGACCGTGCACCTGGACGGCGCTCGATGGTCC 2040
Db 129575 ACGCTTTCGCTCCAAAGATGTCTTCAAGACCGTGCACCTGGACGGCGCTCGATGGTCC 129634
QY 2041 TTTGGGCGCAGCGCGTATTCATTTGCTGCGCGGAGGTTTGTGTGAGGCAATTCG 2100
Db 129635 TTTGGGCGCAGCGCGTATTCATTTGCTGCGCGGAGGTTTGTGTGAGGCAATTCG 129694
QY 2101 CCGAGCGCTTGGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 2160
Db 129695 CCGAGCGCTTGGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 129754
QY 2161 AGAAGTGATTCCTCTCGATGAGATACCCATTCACACCGATGTATGCGAACGATGGCG 2220
Db 129755 AGAAGTGATTCCTCTCGATGAGATACCCATTCACACCGATGTATGCGAACGATGGCG 129814
QY 2221 CTTGATCTAGATCTTAGCTAGTAGACTCACAGCGCGTGTGTGATGACGCAATCGAGG 2280
Db 129815 CTTGATCTAGATCTTAGCTAGTAGACTCACAGCGCGTGTGTGATGACGCAATCGAGG 129874
QY 2281 ATTGGCGCTTACTTCTGAAAAGTTTCAAGGTTTTCACCTTCTTCCCGCGCAGGAA 2340
Db 129875 ATTGGCGCTTACTTCTGAAAAGTTTCAAGGTTTTCACCTTCTTCCCGCGCAGGAA 129934
QY 2341 TTGGGCGCAGCAGAGTAACACCTTTCAGCAATGG 2374
Db 129935 TTGGGCGCAGCAGAGTAACACCTTTCAGCAATGG 129968

RESULT 3

AAF71777

ID AAF71777 standard; DNA; 993 BP.

XX. AAF71777;

AC AAF71777;

XX 30-APR-2001 (first entry)

XX

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.

XX

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;

XX fine chemical production; microorganism; organic acid; nucleoside;

XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX

XX Corynebacterium glutamicum.

XX

XX WO200100843-A2.

XX

XX 04-JAN-2001.

XX

PF 23-JUN-2000; 2000WO-IB00923.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 01-JUL-1999; 99DE-1030476.
 PR 02-JUL-1999; 99US-0142101.
 PR 08-JUL-1999; 99DE-1031415.
 PR 08-JUL-1999; 99DE-1031418.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 08-JUL-1999; 99DE-1031636.
 PR 08-JUL-1999; 99DE-1032125.
 PR 08-JUL-1999; 99DE-1032126.
 PR 08-JUL-1999; 99DE-1032130.
 PR 08-JUL-1999; 99DE-1032186.
 PR 08-JUL-1999; 99DE-1032206.
 PR 08-JUL-1999; 99DE-1032227.
 PR 08-JUL-1999; 99DE-1032228.
 PR 08-JUL-1999; 99DE-1032229.
 PR 08-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;
 XX
 XX WPI; 2001-137957/14.
 DR P-PSDB; AAB79658.
 DR
 XX Nucleic acids from *Corynebacterium glutamicum* encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX

PS Claim 3; Page 226-228; 1737pp; English.
 XX
 CC AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;
 Query Match 41.8%; Score 993; DB 22; Length 993;
 Best Local Similarity 100.0%; Pred. No. 2.6e-299;
 Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1321 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTGA 60
 Db 1 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTGA 60
 Y 1381 AGTAAATGATTGGTCTTAAACATGTTTAAATATAGTCTTCATGAACCCCATTCACCTGA 1440
 Db 61 AGTAAATGATTGGTCTTAAACATGTTTAAATATAGTCTTCATGAACCCCATTCACCTGA 120
 Y 1441 CACTTTGCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCCTTAGCCCTTTCCAT 1500
 Db 121 CACTTTGCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCCTTAGCCCTTTCCAT 180
 Y 1501 TTCCCTCTCGGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTGGGTGAGTGT 1560
 Db 181 TTCCCTCTCGGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTGGGTGAGTGT 240
 Y 1561 GGTATCGCGCACCCACCGCCCAAGACGACCGAGCGGTGAAGTCTTGTGCAAGCAGC 1620
 Db 241 GGTATCGCGCACCCACCGCCCAAGACGACCGAGCGGTGAAGTCTTGTGCAAGCAGC 300
 Y 1621 GCGGAAATGCTGTGCTCAAGCAGAACTAAAGCGCAACTATCTGACGCTTGTCTGA 1680
 Db 301 GCGGAAATGCTGTGCTCAAGCAGAACTAAAGCGCAACTATCTGACGCTTGTCTGA 360
 Y 1681 AATCCGTTAAACATCGCCATCAACGAGATTTCGTATCCACATGTTTCTCCCGTGT 1740
 Db 361 AATCCGTTAAACATCGCCATCAACGAGATTTCGTATCCACATGTTTCTCCCGTGT 420
 Y 1741 CAACGAGGTAGCTTCTTGGGTTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 1800
 Db 421 CAACGAGGTAGCTTCTTGGGTTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 480
 Y 1801 CACATTATCTTGTGCTGCGCGTGGAGATGTTTATAGAGCGGTAAACCGTGAAGCTAATCC 1860
 Db 481 CACATTATCTTGTGCTGCGCGTGGAGATGTTTATAGAGCGGTAAACCGTGAAGCTAATCC 540
 Y 1861 CGTGGCGGATGTAAGTAGTAGAATTTGGAACCATGCGCCACTTGGCCATTTCGAACCC 1920
 Db 541 CGTGGCGGATGTAAGTAGTAGAATTTGGAACCATGCGCCACTTGGCCATTTCGAACCC 600
 Y 1921 CTCATTGGCGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 1980
 Db 601 CTCATTGGCGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 660
 Y 1981 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTCACTGGAAGCGCGCTCGATGGTCC 2040
 Db 661 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTCACTGGAAGCGCGCTCGATGGTCC 720
 Y 2041 TGTGGGCGCAGCGCGTATCCATTGTCCGTCGCGGAAGGTTTGTGTGAGGCAATTGG 2100
 Db 721 TGTGGGCGCAGCGCGTATCCATTGTCCGTCGCGGAAGGTTTGTGTGAGGCAATTGG 780
 Y 2101 CCGAGGCTTGTGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 2160
 Db 781 CCGAGGCTTGTGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 840
 Y 2161 AGAAGTGTATCTCTCGATGAGATACCCATTGACACCGGATGATTGGCAACGATGGG 2220

```

Db 841 AGAAGTATCTCTCTGATGAGATACCAATGACACACCGATGATATGGCAAGTGGCG 900
QY 2221 CCTGGAATCTAGATCTCTAGCTAGACTCAGAGCGCGCTGTTGATGACGAATCGAGGG 2280
Db 901 CCTGGAATCTAGATCTCTAGCTAGACTCAGAGCGCGCTGTTGATGACGAATCGAGGG 960
QY 2281 ATTGGCGCCTTAGTACTCTCTGAAAAGTTTCAG 2313
Db 961 ATTGGCGCCTTAGTACTCTCTGAAAAGTTTCAG 993

RESULT 4
AAS96096
ID AAS96096 standard; DNA; 993 BP.
XX
AC AAS96096;
XX
DT 26-FEB-2002 (first entry)
XX
DE C. glutamicum gene #21 encoding metabolic pathway protein.
XX
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
KW methionine biosynthesis pathway; large-scale production of fine chemical;
KW Corynebacterium diphtheriae; diphtheria; ds.
XX
OS Corynebacterium glutamicum.
XX
PN W0200166573-A2.
XX
PD 13-SEP-2001.
XX
PF 22-DEC-2000; 2000WO-1B02035.
XX
PR 09-MAR-2000; 2000US-187970P.
PR 23-JUN-2000; 2000US-0606740.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
PI Lee H, Hwang B;
XX
DR WPI: 2001-582269/65.
DR P-PSDB; AAU71886.
XX
XX
PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium
PT glutamicum, useful for producing methionine and lysine in
PT Corynebacterium and Brevibacterium.
XX
PS Disclosure; Page 207-208; 316pp; English.
XX
CC The present invention relates to the isolation of novel Corynebacterium
CC glutamicum genes encoding metabolic pathway (MP) proteins
CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention
CC include enzymes involved in the lysine and methionine biosynthetic
CC pathways. The polynucleotide sequences of the invention can be used
CC for the large-scale production and/or modulation of expression of the
CC fine chemicals such as lysine and methionine. The sequences of the
CC invention may be used to identify C. glutamicum and related organisms
CC e.g. C. diphtheriae in a subject to detect diphtheria.
CC AAS96073-AAS96132 represent C. glutamicum genes encoding the novel
CC metabolic pathway proteins of the invention.
XX
SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match 41.8%; Score 993; DB 23; Length 993;
Best Local Similarity 100.0%; Pred. No. 2.6e-299;
Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1321 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
Db 1 CAAAAGCAGACCTGTAATGAAGATTTCATGATCACCATCGTGACCTATGGAAGTACTTA 60

```

```

QY 1381 AGTAAATGATGGTCTTAAACATGGTTTAAATATAGCTTCATGAACCCCAATTCACATCGA 1440
Db 61 AGTAAATGATGGTCTTAAACATGGTTTAAATATAGCTTCATGAACCCCAATTCACATCGA 120
QY 1441 CACTTGTCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCTTAGCCCTTTCCAT 1500
Db 121 CACTTGTCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCTTAGCCCTTTCCAT 180
QY 1501 TTCCCTCTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
Db 181 TTCCCTCTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 240
QY 1561 GGTATCGCGCACCCCAACCGGCCAAAGCAACCGAGGGGTGAAGTCTTGTGCAAGCAGC 1620
Db 241 GGTATCGCGCACCCCAACCGGCCAAAGCAACCGAGGGGTGAAGTCTTGTGCAAGCAGC 300
QY 1621 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTGTCTGA 1680
Db 301 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTGTCTGA 360
QY 1681 AATCCGTTAAACATCGCCATCAACGAGATTCGCTATCCATGTTTCTCCCGTGT 1740
Db 361 AATCCGTTAAACATCGCCATCAACGAGATTCGCTATCCATGTTTCTCCCGTGT 420
QY 1741 CAACGAGGTAGTCTTCTGGGTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 1800
Db 421 CAACGAGGTAGTCTTCTGGGTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 480
QY 1801 CACATTATCTTCTGCGCGGTGGAGATGTTTATAGAGCGGTAAACCGGTGAAGCTAATCC 1860
Db 481 CACATTATCTTCTGCGCGGTGGAGATGTTTATAGAGCGGTAAACCGGTGAAGCTAATCC 540
QY 1861 CGTGGCGGATGTAAGTAGTAGAACTTGGAAACCATGCGCCACTTGGCCATTCGAACCCC 1920
Db 541 CGTGGCGGATGTAAGTAGTAGAACTTGGAAACCATGCGCCACTTGGCCATTCGAACCCC 600
QY 1921 CTCATTGCGGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGTCTT 1980
Db 601 CTCATTGCGGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGTCTT 660
QY 1981 AGCTTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGAGCGGCGCGTGCATGGTCC 2040
Db 661 AGCTTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGAGCGGCGCGTGCATGGTCC 720
QY 2041 TGTGGGCGCAGCGCGTATCCATTGTCCTGCGCGGAAGGTTTTGTGAGGCAATTCG 2100
Db 721 TGTGGGCGCAGCGCGTATCCATTGTCCTGCGCGGAAGGTTTTGTGAGGCAATTCG 780
QY 2101 CCGAGGCTTGGTGGGAGCTTCTCCGAAACCCAGCTGCTCCCATGCTAAAGCAGG 2160
Db 781 CCGAGGCTTGGTGGGAGCTTCTCCGAAACCCAGCTGCTCCCATGCTAAAGCAGG 840
QY 2161 AGAAGTGATCTCTCTCGATGAGATACCATGACACACCGATGATATGGCAACGATGGCG 2220
Db 841 AGAAGTGATCTCTCTCGATGAGATACCATGACACACCGATGATATGGCAACGATGGCG 900
QY 2221 CTGGAATCTAGATCTCTAGCTAGACTCAGAGCGCGCTGTTGATGCAAGCAATCGAGG 2280
Db 901 CTGGAATCTAGATCTCTAGCTAGACTCAGAGCGCGCTGTTGATGCAAGCAATCGAGG 960
QY 2281 ATTGGCGCCTTAGTACTCTCTGAAAAGTTTCAG 2313
Db 961 ATTGGCGCCTTAGTACTCTCTGAAAAGTTTCAG 993

RESULT 5
AAH68421
ID AAH68421 standard; DNA; 870 BP.
XX
AC AAH68421;
XX
DT 26-SEP-2001 (first entry)
XX

```


C glutamicum coding sequence fragment SEQ ID NO: 3456.

Corynebacterium; amino acid synthesis; vitamin; saccharide;
organic acid synthesis; ds.

Corynebacterium glutamicum.

EP1108790-A2.

20-JUN-2001.

18-DEC-2000; 2000EP-0127698.

16-DEC-1999; 99JP-0377484.

07-APR-2000; 2000JP-0159162.

03-AUG-2000; 2000JP-0280988.

(KYOWA) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI: 2001-376931/40.

P-PSDB; AAG93202.

Novel polynucleotides derived from Corynebacterium bacteria, for identifying
mutation point of a gene, measuring expression of a gene, analysing
expression profile or pattern of a gene and identifying homologous gene

Claim 1; SEQ ID NO: 3456; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein
sequences from the Corynebacterium glutamicum. These
are useful for identifying the mutation point of a gene derived from a
mutant of corynebacterium, measuring expression amount and
analysing the expression profile or expression pattern of a gene derived
from Corynebacterium, and identifying a homologue of a gene derived
from corynebacterium. Corynebacterium bacteria are useful for producing
amino acids, nucleic acids, vitamins, saccharides and organic acids,
particularly L-lysine. The present sequence is a nucleic acid described
in the exemplification of the invention.

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from the
European Patent Office.

Sequence 870 BP; 192 A; 236 C; 245 G; 197 T; 0 other;

Query Match 36.6%; Score 870; DB 22; Length 870;
Best Local Similarity 100.0%; Pred. No. 6.6e-261;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1421 ATGACCCCACTTCACTGGACCTTCTCTCAATCATTTGATGAGCAGCTTGAAGGC 1480

1 ATGACCCCACTTCACTGGACCTTCTCTCAATCATTTGATGAGCAGCTTGAAGGC 60

QY 1481 GCCTCTTAGCCCTTTCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 1540

61 GCCTCTTAGCCCTTTCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 120

QY 1541 CATCAGTGGGTGAGTGTGTTGATTCGCGCACCCCAAGCAAGCAAGCGGGT 1600

121 CATCAGTGGGTGAGTGTGTTGATTCGCGCACCCCAAGCAAGCAAGCGGGT 180

QY 1601 GAAGTCTTGTGACAGCGCGGAAATGGTGTGTCGACAGCAAACTAAAGCGCAA 1660

181 GAAGTCTTGTGACAGCGCGGAAATGGTGTGTCGACAGCAAACTAAAGCGCAA 240

QY 1661 CTATCTGAGCGCTTGTGTAATCCCGTTAAACCATCCCATCAACGAGATTCGCTATCC 1720

241 CTATCTGAGCGCTTGTGTAATCCCGTTAAACCATCCCATCAACGAGATTCGCTATCC 300

QY 1721 ACANGTTCCTCCCGTGTTCACAGAGGTAGTCTTCTTGGGGTGGAGCAAGCTCAAGCTG 1780

Db 301 ACATGGTTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGTGGACCAAGCTCACGCTG 360

QY 1781 CGTTTGGAGATGAAGCGCACATTTATCTTCTCGCGGTGGAGATGTTTGGAGCG 1840

Db 361 CGTTTGGAGATGAAGCGCACATTTATCTTCTCGCGGTGGAGATGTTTGGAGCG 420

QY 1841 GTAAACCCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGTGAACCTTGAACCATGCGC 1900

Db 421 GTAAACCCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGTGAACCATGCGC 480

QY 1901 CACTTGGCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGGAACTAGAT 1960

Db 481 CACTTGGCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGGAACTAGAT 540

QY 1961 TGGGCTGCGATGCCCTCTTACGCTTCGGTCCCAAGATGTGCTCAAGACCGTGACCTG 2020

Db 541 TGGGCTGCGATGCCCTCTTACGCTTCGGTCCCAAGATGTGCTCAAGACCGTGACCTG 600

QY 2021 GACGGCGCGTTCGATGCTCTGTGGGGCGCAGCGGTATCCATTGCTCCGTCGCGCGAA 2080

Db 601 GACGGCGCGTTCGATGCTCTGTGGGGCGCAGCGGTATCCATTGCTCCGTCGCGCGAA 660

QY 2081 GGTTCGTCGAGCAATTCGCGAGGCTTGTGTTGGGGACTTCTTCCCGAAACCCCAAGCT 2140

Db 661 GGTTCGTCGAGCAATTCGCGAGGCTTGTGTTGGGGACTTCTTCCCGAAACCCCAAGCT 720

QY 2141 GCTCCCATGCTAAAGCAGAGAGTGTGCTTCTCGATGAGATACCCATTGACACACCG 2200

Db 721 GCTCCCATGCTAAAGCAGAGAGTGTGCTTCTCGATGAGATACCCATTGACACACCG 780

QY 2201 ATGTATTGCAAGCGGCGCTTGGAACTAGATCTCTAGTAGTCTACAGACCGCTC 2260

Db 781 ATGTATTGCAAGCGGCGCTTGGAACTAGATCTCTAGTAGTCTACAGACCGCTC 840

QY 2261 GTTGTGTCGAGCAATTCGAGGATTCGCGCT 2290

Db 841 GTTGTGTCGAGCAATTCGAGGATTCGCGCT 870

RESULT 6

AAF71779/c

ID AAF71779 standard; DNA; 822 BP.

XX AAF71779;

AC AAF71779;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.

XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;

XX fine chemical production; microorganism; organic acid; nucleoside;

XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

XX WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

PR 01-JUL-1999; 99DE-1030476.

PR 02-JUL-1999; 99US-0142101.

PR 08-JUL-1999; 99DE-1031415.

PR 08-JUL-1999; 99DE-1031418.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031435.
 PR 08-JUL-1999; 99DE-1031443.
 PR 08-JUL-1999; 99DE-1031453.
 PR 08-JUL-1999; 99DE-1031457.
 PR 08-JUL-1999; 99DE-1031465.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031510.
 PR 08-JUL-1999; 99DE-1031541.
 PR 08-JUL-1999; 99DE-1031573.
 PR 08-JUL-1999; 99DE-1031592.
 PR 08-JUL-1999; 99DE-1031632.
 PR 08-JUL-1999; 99DE-1031634.
 PR 08-JUL-1999; 99DE-1031636.
 PR 08-JUL-1999; 99DE-1032125.
 PR 08-JUL-1999; 99DE-1032126.
 PR 08-JUL-1999; 99DE-1032130.
 PR 08-JUL-1999; 99DE-1032186.
 PR 08-JUL-1999; 99DE-1032206.
 PR 08-JUL-1999; 99DE-1032227.
 PR 08-JUL-1999; 99DE-1032228.
 PR 08-JUL-1999; 99DE-1032229.
 PR 08-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032926.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1033004.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 12-AUG-1999; 99US-0148613.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040832.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041380.
 PR 31-AUG-1999; 99DE-1041394.
 PR 31-AUG-1999; 99DE-1041396.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042124.
 PR 03-SEP-1999; 99DE-1042129.
 PR 09-MAR-2000; 2000US-0187970.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 XX
 XX WPI: 2001-137957/14.
 DR P-PSDB; AAB79660.
 DR
 XX Nucleic acids from *Corynebacterium glutamicum* encoding metabolic
 PT pathway proteins, useful for producing fine chemicals in
 PT microorganisms, including organic acids, nonproteinogenic amino acids,
 PT and purine and pyrimidine bases -
 XX
 XX Claim 3; Page 233-234; 1737pp; English.
 XX
 CC AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The *C. glutamicum*
 CC MP nucleic acids are useful for the production of fine chemicals
 CC in microorganisms, including organic acids, nonproteinogenic amino
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
 CC compounds, vitamins, cofactors, polyketides and enzymes.
 XX
 SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 34.6%; Score 822; DB 22; Length 822;
 Best Local Similarity 100.0%; Pred. No. 6.5e-246;
 Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 629 GATTCACAAACCGCGAAACTAACCCATCAACATCAGTTTTCATGGCCCAATCGCGTCATC 688
 DB 822 GATTCACAAACCGCGAAACTAACCCATCAACATCAGTTTTCATGGCCCAATCGCGTCATC 763
 QY 689 ACAACTGCCACGACGAGTGTGATCCAGGCCCAACCTTGGGGCTGGACAGCGGGCGGTGAC 748
 DB 762 ACAACTGCCACGACGAGTGTGATCCAGGCCCAACCTTGGGGCTGGACAGCGGGCGGTGAC 703
 QY 749 AATGCTGCTGCCGCGAAACCCAGCGGGAACCCAGATCAGGCTTCCGCGCAACCGGCCA 808
 DB 702 AATGCTGCTGCCGCGAAACCCAGCGGGAACCCAGATCAGGCTTCCGCGCAACCGGCCA 643
 QY 809 GGGGCGAAATCCACCGTCCGGTGTTCGGCGTATTGGCGCCGACGCGCGCCGATAAACACA 868
 DB 642 GGGGCGAAATCCACCGTCCGGTGTTCGGCGTATTGGCGCCGACGCGCGCCGATAAACACA 583
 QY 869 AACGGGTCCAAATACGCAATTCCGGGTTCACACGAGTCCAGCAGATTCGCCATCAACATGGGC 928
 DB 582 AACGGGTCCAAATACGCAATTCCGGGTTCACACGAGTCCAGCAGATTCGCCATCAACATGGGC 523
 QY 929 TTATACCAAAACCGCTGCTTATCGAGCTCAGCTCCACCGCACCGCGTGGCGGTGTCAC 988
 DB 522 TTATACCAAAACCGCTGCTTATCGAGCTCAGCTCCACCGCACCGCGTGGCGGTGTCAC 463
 QY 989 GTGGCCACCGCGCAACCGCGCGTGTTCATCGGGCAGCGTGTTCGTTCTGTTCTTCA 1048
 DB 462 GTGGCCACCGCGCAACCGCGCGTGTTCATCGGGCAGCGTGTTCGTTCTGTTCTTCA 403
 QY 1049 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGTCTTCGCTGCCATGACGCAAC 1108
 DB 402 ATGATCTGTGGCGCTTCCACCTTGTTCATGGCGTCTTCGCTGCCATGACGCAAC 343
 QY 1109 CATAACAGGTAAGCGATGCGACCCCGCGCATATATCGAGCAGATCGCGCGCGCATTTG 1168
 DB 342 CATAACAGGTAAGCGATGCGACCCCGCGCATATATCGAGCAGATCGCGCGCGCATTTG 283
 QY 1169 GACAAAAGATCAAGCGCCCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 1228
 DB 282 GACAAAAGATCAAGCGCCCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 223
 QY 1229 ACAGAGAACCGCAATGAGTCTTCGCGCTTAATTCCTTGTTCATACCAAGTACATTC 1288
 DB 222 ACAGAGAACCGCAATGAGTCTTCGCGCTTAATTCCTTGTTCATACCAAGTACATTC 163
 QY 1289 TCGGCTCCGATGGACAGTAAGAGCTGGCGCCCGCAAGACGACCTGTATGAGATTTC 1348
 DB 162 TCGGCTCCGATGGACAGTAAGAGCTGGCGCCCGCAAGACGACCTGTATGAGATTTC 103
 QY 1349 ATGATCACCATCGTGACCTATGAGTACTTAAGTAAATGATTGGTCTTAAACATGGTT 1408
 DB 102 ATGATCACCATCGTGACCTATGAGTACTTAAGTAAATGATTGGTCTTAAACATGGTT 43
 QY 1409 TAATATAGCTTCATGAACCCCATTTCACTGGACACTTTGCTC 1450
 DB 42 TAATATAGCTTCATGAACCCCATTTCACTGGACACTTTGCTC 1

RESULT 7

AAS96098/c

ID AAS96098 standard; DNA; 822 BP.

XX AAS96098;

XX 26-FEB-2002 (first entry)

XX C. glutamicum gene #23 encoding metabolic pathway protein.

XX Metabolic pathway protein; MP; lysine biosynthesis pathway;

methionine biosynthesis pathway; large-scale production of fine chemical;
Corynebacterium diphtheriae; diphtheria; ds.
Corynebacterium glutamicum.

QY 1049 ATGATCTGTGGCGCTTCCACCTTGTTCATGCGGTCTTTGCTGCCATGACGCGCAAC 1108
DB 402 ATGATCTGTGGCGCTTCCACCTTGTTCATGCGGTCTTTGCTGCCATGACGCGCAAC 343
QY 1109 CATAACAGGTAAGCGATGCGCACCAGCGCATATAATCGAGCAGCATCGCGCGGCAATTG 1168
DB 342 CATAACAGGTAAGCGATGCGCACCAGCGCATATAATCGAGCAGCATCGCGCGGCAATTG 283
QY 1169 GACAAAAGATCAACGCCCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 1228
DB 282 GACAAAAGATCAACGCCCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 223
QY 1229 ACGAGAAGAACCGCAATGAGTCTTCCGCTTAAATTCCTTGTTCATCAGCATATTC 1288
DB 222 ACGAGAAGAACCGCAATGAGTCTTCCGCTTAAATTCCTTGTTCATCAGCATATTC 163
QY 1289 TGGCGTCCGATGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTAATGAAGATTTC 1348
DB 162 TGGCGTCCGATGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTAATGAAGATTTC 103
QY 1349 ATGATCACCATCGTGAACCTATGGAAGTACTTAAGTAAATGATTGTTCTTAACATGGTT 1408
DB 102 ATGATCACCATCGTGAACCTATGGAAGTACTTAAGTAAATGATTGTTCTTAACATGGTT 43
QY 1409 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTGCTC 1450
DB 42 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTGCTC 1
RESULT 8
AAH68420/c
ID AAH68420 standard; DNA; 708 BP.
AC AAH68420;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 3455.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG93201.
XX
DR Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 1; SEQ ID NO: 3455; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum.
CC These are useful for identifying the mutation point of a gene derived from a

Query Match 34.6%; Score 822; DB 23; Length 822;
Best Local Similarity 100.0%; Pred. No. 6.5e-246;
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCCAAACCCGGAACCTAACCCATCAACATCAGTTGATGGCCATGCGGTATC 688
DB 822 GATTCCAAACCCGGAACCTAACCCATCAACATCAGTTGATGGCCATGCGGTATC 763
QY 689 ACAACTGCCAGCAGCGATTTGATCCAGCGCACACCTTGGGGCTGGACAGCGCGGTGAC 748
DB 762 ACAACTGCCAGCAGCGATTTGATCCAGCGCACACCTTGGGGCTGGACAGCGCGGTGAC 703
QY 749 AATGCTGCTGGCGGAACCCACAGCGGAACAGATCAGGCTTGGCGGAACCGGCCA 808
DB 702 AATGCTGCTGGCGGAACCCACAGCGGAACAGATCAGGCTTGGCGGAACCGGCCA 643
QY 809 GCGCGGAACCTCACCGTCCGCTGTCGCGGTATGCGCGCGAGCGCGCGATTAACACA 868
DB 642 GCGCGGAACCTCACCGTCCGCTGTCGCGGTATGCGCGCGAGCGCGCGATTAACACA 583
QY 869 AAGCGGTCCAAATACGATTCGGGTTCACACAGGTGACAGCATGCGCATCAACATGGGC 928
DB 582 AAGCGGTCCAAATACGATTCGGGTTCACACAGGTGACAGCATGCGCATCAACATGGGC 523
QY 929 TTTACCAAAACCCGCTGTTATCGAGCTCACCTCCACCGCACCGGTGCGCGGTGCA 988
DB 522 TTTACCAAAACCCGCTGTTATCGAGCTCACCTCCACCGCACCGGTGCGCGGTGCA 463
QY 989 GTGGCCACCGCCAAACCGCCAAAGGGGTGTCATCGGGCAACGTTGTTCTTCA 1048
DB 462 GTGGCCACCGCCAAACCGCCAAAGGGGTGTCATCGGGCAACGTTGTTCTTCA 403

QY 1190 GTCCGGCGATGAACAAAGAGACGCTCAGAAATTAACACACAGAGAAACCGCAATAGT 1249
| | | | |
Db 947 GTGCCGAGGGTGAACAGGACACGACGCTCGACAGCAGACAGATGATGACGGCGTGTATG 888
| | | | |
QY 1250 CTTTCGGCTTAATCTCTGTTTAAATCACCAGTACATCTCGCGTCCGATGAGCAAGTAAA 1309
| | | | |
Db 887 CCCTCGGGTTGATGCGCTGTTGATCACCAGGACATCTGTGGGCGCGATGGCCAAACAGC 828
| | | | |
QY 1310 AGACTGGCCCCCAAGAGCAGACCTGTATGAAGATTTCCATGATCACCATCGTGACCTAT 1369
| | | | |
Db 827 AGACTGGCTCCCAACACAAACCGGTGACAAAGATTTCCATGTCCCGATGTTCCGCCAC 768
| | | | |
QY 1370 GGAAGTACTTAAGTAAATGATGGTCTTAATGATGGTTTAAATAGCTTCATGAACCCC 1429
| | | | |
Db 767 AAACCAATTAAGTAAATGATGAGACTGAACATGGTTTGAATGCTTCATGAACCCG 708
| | | | |
QY 1430 ATTCAATGAGACCTTGTCTCTCAATCATTTGATGAAGGAGCTTCGAAGGGCCCTCTTA 1489
| | | | |
Db 707 ATCCACTGGACACCTTCTGACCATCATCGACGAGGCGAGCTTCGAGAAACGCCCTCCCTC 648
| | | | |
QY 1490 GCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTG 1549
| | | | |
Db 647 GCATGTGATCTCCCTCGCGGTGAGTCAGCGGCTCAAGGCACCTGGAGAAATCCGTC 588
| | | | |
QY 1550 GGTGAGTGTGTGATCGCGCACCAACCGGCCAAAGCGGGGTGAAGTCTCTT 1609
| | | | |
Db 587 GGTGCGGTACTGTGTGCGCGCACCGCGGTTCGCCAAGCGGGGAGTGTCTG 528
| | | | |
QY 1610 GTGCAAGCAGCGGAAATGTTGTTCTGTCAGCAGCAAACTAAAGCGCAACTATCTGGA 1669
| | | | |
Db 527 GTCCAGCGCGCCCAAGATGGCGCTCTGCGAGCGGAGACCCGTGAGCAACTGGCGGAA 468
| | | | |
QY 1670 CGCTTCGCTGAATCCCGTTAAACCATCCATCAACGACGATTCGCTATCCACATGTTT 1729
| | | | |
Db 467 CGGTTGACGAGATCCCTCGACCGTGTGCTATCAACGCGATTCCTTGTCCACTGTTTC 408
| | | | |
QY 1730 CTCTCCGTTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTCGCTTGGAA 1789
| | | | |
Db 407 CCGCCCGTGTTCGCGAGGTGTCACATTTGGGGTGGCTGACCTGACGCTGGGTGGAG 348
| | | | |
QY 1790 GATGAAGCGCACATATCTCTGCTGGGGTGGAGATGTTTGAAGCGGTAAACCGT 1849
| | | | |
Db 347 GACGAGCCCAACACCTGTCTCTGTTGCGCGGTTCCGCTCGGGGCGCGTACCCCGC 288
| | | | |
QY 1850 GAAGCTAATCCCGTGGCGGATGTAAGTAGTAACTTGAACCACTGCGCCACTTGGCC 1909
| | | | |
Db 287 GAGCGGACCCGTTGCGGGTGTAGGTCTACGCTCGGGGTATGAGACACCTGCGG 228
| | | | |
QY 1910 ATTGCAACCCCTCTCATTTGCGGGATGCTTACATGTTGATGGAAACTAGATTGGGCTGCG 1969
| | | | |
Db 227 GTGGCCACCCCGGAACCTGAGGCAAGCTACACGTTGATGGCCACACCGACTTGGGTGCG 168
| | | | |
QY 1970 -ATGCCGCTTACGCTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGACGCGGCG 2028
| | | | |
Db 167 CATGCGGCTACTGCTTTCGGCCGCAATGATGTCTGCTGACGACCGCGATCTCGAGGCGAG 108
| | | | |
QY 2029 CGTGCATGTCTGTTGGGGCGGAGCGCTATCCATTTGCTCCGCTCGCGGCAAGGTTTGG 2088
| | | | |
Db 107 GGTGACGAGACCGTCCCAAGGCGCGGTGTCTCGGTGTGCGGTGCGGAGGTTTCGG 48
| | | | |
QY 2089 TGAGGCAATTTGCGCGAGGCTTGTGTTGGGACTTCTTCCCGAAACCC 2135
| | | | |
Db 47 GGAGGCGGTGCGACTCGGCTCGGGTGGGCGCTGCTGCCCGAGGCC 1

RESULT 10

AAH68419

ID AAH68419 standard; DNA; 627 BP.

XX

AC AAH68419;

XX

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 3454.
DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
KW Corynebacterium glutamicum.
XX EP1108790-A2.
XX 20-JUN-2001.
XX 18-DEC-2000; 2000EP-0127688.
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
DR P-PSDB; AAG93200.
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 1; SEQ ID NO: 3454; 246bp + Sequence Listing; English.
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

Query Match 26.4%; Score 627; DB 22; Length 627;

Best Local Similarity 100.0%; Pred. No. 5e-185;

Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 ATGTACGCAATTCGCTGACATTTGTCGTCTGGAAGGCTCTTTAGTGGTATTTCTTCC 82
| | | | |
Db 1 ATGTACGCAATTCGCTGACATTTGTCGTCTGGAAGGCTCTTTAGTGGTATTTCTTCC 60
| | | | |
QY 83 TACGCTCCAGAGCTCAGCGGAGCGGCTGAGTTCATGCGGAGAGAGGCTGCGCGT 142
| | | | |
Db 61 TACGCTCCAGAGCTCAGCGGAGCGGCTGAGTTCATGCGGAGAGAGGCTGCGCGT 120
| | | | |
QY 143 CTGATTTCATCAGCAAGCTATTCATTAATCTGTTGGTGGAGGACCGGCGATGAC 202
| | | | |
Db 121 CTGATTTCATCAGCAAGCTATTCATTAATCTGTTGGTGGAGGACCGGCGATGAC 180
| | | | |
QY 203 GGTGAGAACTTGTTCAGTCAGTCAGTCCCAACATGCTTGGCGTCAATGCTTTCTCACC 262
| | | | |
Db 181 GGTGAGAACTTGTTCAGTCAGTCAGTCCCAACATGCTTGGCGTCAATGCTTTCTCACC 240
| | | | |
QY 263 CTTGCGAGGCGCTGCTCAGGACAAATATCTCGATGGAATTCAGAGGTTCCCGCGCC 322
| | | | |
Db 241 CTTGCGAGGCGCTGCTCAGGACAAATATCTCGATGGAATTCAGAGGTTCCCGCGCC 300
| | | | |

QY 323 AGCCAGGTAAGTCCCTGCTCTGAGGGCATGTTGAACGTGAACAATATTGATATGTCGCGC 382
 DB 301 AGCCAGGTAAGTCCCTGCTCTGAGGGCATGTTGAACGTGAACAATATTGATATGTCGCGC 360
 QY 383 AAGCTCAATCAGATCGCCAGGAAACGGGGCAGTCACTTCGCGAGATGGCGCTTGCATGG 442
 DB 361 AAGCTCAATCAGATCGCCAGGAAACGGGGCAGTCACTTCGCGAGATGGCGCTTGCATGG 420
 QY 443 GTGCTGCGGAGCAAGAGAGATACGGCGCGATACCGTGACCAAGTGCATTTGATGCT 502
 DB 421 GTGCTGCGGAGCAAGAGAGATACGGCGCGATACCGTGACCAAGTGCATTTGATGCT 480
 QY 503 TCGTCAGTTGAGCAGCTGGCAACAGCGCTTGATTCATCAACACTTGGATTTCTGAC 562
 DB 481 TCGTCAGTTGAGCAGCTGGCAACAGCGCTTGATTCATCAACACTTGGATTTCTGAC 540
 QY 563 GCCGAGTTGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGCGGAAG 622
 DB 541 GCCGAGTTGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGCGGAAG 600
 QY 623 GCCACGATTCGAAACCCGCGAAAC 649
 DB 601 GCCACGATTCGAAACCCGCGAAAC 627

RESULT 11

AAF68077
 ID AAF68077 standard; DNA; 993 BP.
 AC AAF68077;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:669.
 XX
 KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
 KW membrane construction and membrane transport protein; petroleum spill;
 KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
 KW identification; microorganism; fine chemical production; transformation;
 KW genome mapping; genetic engineering; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 XX WC200100805-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00926.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031563.
 PR 08-JUL-1999; 99DE-1032122.
 PR 08-JUL-1999; 99DE-1032124.
 PR 08-JUL-1999; 99DE-1032125.
 PR 08-JUL-1999; 99DE-1032128.
 PR 08-JUL-1999; 99DE-1032180.
 PR 08-JUL-1999; 99DE-1032182.
 PR 08-JUL-1999; 99DE-1032190.
 PR 08-JUL-1999; 99DE-1032191.
 PR 08-JUL-1999; 99DE-1032209.
 PR 08-JUL-1999; 99DE-1032212.
 PR 08-JUL-1999; 99DE-1032227.
 PR 08-JUL-1999; 99DE-1032228.
 PR 08-JUL-1999; 99DE-1032229.
 PR 08-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.

PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041395.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX (BADI) BASF AG.
 PA
 XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;
 PI
 XX WPI; 2001-071486/08.
 DR P-PSDB; AAB76844.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation -
 XX
 XX Claim 3; Page 1107-1108; 1119pp; English.
 PS
 CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 XX
 SQ Sequence 993 BP; 222 A; 247 C; 277 G; 247 T; 0 other;
 Query Match 20.4%; Score 485.2; DB 22; Length 993;
 Best Local Similarity 99.0%; Pred. No. 1.4e-140;
 Matches 499; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
 QY 1 AGATACCTCTTTGGAGAAACCATGTACGCATTCGCTGACATTTGCTCTGGAAGGC 60
 DB 490 AGATACCTCTTTGGAGAAACCATGTACGCATTCGCTGACATTTGCTCTGGAAGGC 549
 QY 61 TCTTTACGTGGGTATTTCTTCTACGCTCCAGAGCTCACAGCGAGGGCGGTGATTCA 120
 DB 550 TCTTTACGTGGGTATTTCTTCTACGCTCCAGAGCTCACAGCGAGGGCGGTGATTCA 609
 QY 121 GCGGAGGAGGCTGCGCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATCGTTG 180
 DB 610 GCGGAGGAGGCTGCGCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATCGTTG 669
 QY 181 GGTGGAGGAACCGGGCGATGACGGTGAGAACTTTGTCAGTCAGTCCCAATGGTCT 240
 DB 670 GGTGGAGGAACCGGGCGATGACGGTGAGAACTTTGTCAGTCAGTCCCAATGGTCT 729
 QY 241 TGGCGTCATTTGCTTCTCACCACCTTCGCGAGGGCTGCTCAGGACAAATATCTCGATGG 300
 DB 730 TGGCGTCATTTGCTTCTCACCACCTTCGCGAGGGCTGCTCAGGACAAATATCTCGATGG 789
 QY 301 AATTCCAGAGGGTTCGCGCCAGCGAGGTAAGTCCCTGTCTGAGGGGATGTTGAACGT 360
 DB 790 AATTCCAGAGGGTTCGCGCCAGCGAGGTAAGTCCCTGTCTGAGGGGATGTTGAACGT 849
 QY 361 GAACAATATTGATGCTCGCAAGCTCAATGACATCGCCCGAGGAACGCGGGCAGTCACT 420
 DB 850 GAACAATATTGATGCTCGCAAGCTCAATGACATCGCCCGAGGAACGCGGGCAGTCACT 909
 QY 421 TGGCGCAGATGGCGCTTGTGATGGTGTCTGCGCAGCAAGGAGAGTACGGCGC--GGATACC 478

Db 910 TGCCAGATGCGCTGCGTGGTGTGCGCGAGCAAGAGAGTACGCGCGCGATTACC 969
 Qy 479 GTGACCACTGCTGATGATGATGCT 502
 Db 970 GTGACCACTGCTGATGATGCT 993

RESULT 12

AA199683/c

ID AAT33536 standard; DNA; 15239 BP.

AC AAT33536;

XX 15-FEB-1998 (first entry)

XX BCG deletion region 2 and flanking sequences.

XX BCG delta 2; virulence; avirulence; attenuation; gene deletion;

XX Mycobacteria; vaccine; infection; marker; ss.

XX Mycobacterium bovis strain BCG.

FH Key Location/Qualifiers

FT misc_feature 3382..14071

FT /*tag= a

FT /note= "BCG delta 1 deletion region"

XX W09625519-A1.

XX 22-AUG-1996.

XX 15-FEB-1996; 96WO-US01938.

XX 17-FEB-1995; 95US-0390878.

XX (PATH-) PATHOGENESIS CORP.

XX Mahairas GG, Stover CK;

XX WPI; 1996-393419/39.

XX Detecting markers for avirulence in Mycobacterium - used in

XX production of vaccines against bacterial infection, and to detect

XX bacterial infection

XX Example 1; Fig 2; 66pp; English.

XX This DNA sequence comprises Mycobacterium bovis BCG deletion

XX sequence BCGdelta2. A specific genetic deletion of this region

XX results in an avirulence phenotype of the mycobacterium. 2 Other

XX deletion regions (see AAT33535 and AAT33537) have also been detected.

XX Identification involved screening a BCG cosmid library with a

XX radiolabeled probe obtained following DNA subtraction between

XX virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.

XX The deletions provide useful markers for the identification of an

XX avirulent, or a virulent, mycobacterial phenotype. Determination

XX of avirulence requires the detection of the presence or absence of

XX the deletion; the deletions are detected either by detecting the

XX presence or absence of deletion junctions (see AAT33538-46), or by

XX detecting the presence or absence of the sequences contained within

XX the deletion. Deletion polypeptides are used as components of

XX immunological assays and in vaccines.

XX Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

XX Query Match 5.2%; Score 122.8; DB 17; Length 15239;

XX Best Local Similarity 49.7%; Pred. No. 1.5e-26;

XX Matches 434; Conservative 0; Mismatches 422; Indels 18; Gaps 4;

Qy 1433 CAACTGACACTTGTCTCAATCATGATGAAGGAGCTTCGAAGCGGCTCTTAGCG 1492

Db 5253 CAGCTGGCCGCAATTGGTCCGCGGTGCGAAGTCTGATGCGCGCGGAGCGC 5312

Qy 1493 CTTTCCATTTCCCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACGTTGGGT 1552
 Db 5313 CTACATGTCACTCCCTCGCTGCTGCTGAGCGCATCACTGCTGTTGAGCAGGAGTCCGC 5372
 Qy 1553 CGAGTGTGTTATCGCGCACCCCAACCGGCCAAAGCAACCGAGCGGTGAAGTCTTGTG 1612
 Db 5373 CAGTGTCTGTTGTCAGGGGAAAGCCATGTCGGGCGCAGCCAGGTATCCGCTGTTG 5432
 Qy 1613 CAAGCAGCGCGGAAATGTTGTCAGCAGCAAACTAAAGCGCAACTATCTGG--AC 1670
 Db 5433 CGTTGCGCGCGCAACACAGCGTTGCTCGAGTCCGAGCGCTCGCTGAAATGGTGGCAAC 5492
 Qy 1671 GCCTTGCTGAA-----ATCCCGTTAAACATCGCCATCAACCGCAGATTCGCTATCCATCG 1726
 Db 5493 GCGTGTCTGAACGCGCAGCGGATCACCATTGCGGTAAAGCGCGATTCATGCGGACATGG 5552
 Qy 1727 TTTCTCCCGTGTTCACAGAGGTAGCTTCTGGGGTGGAGCAACGCTCAGCTCGCGCTTG 1786
 Db 5553 TTTTCCGCGCTGTTCCGACG-----GTCTCGGCGACGCTCTGCTCGACGTTGGGATC 5603
 Qy 1787 GAAGATGAAGCGCACACATTTATCTTCTGCTGCGCGGTGGAGATGTTTTAGGAGCGGTAAAC 1846
 Db 5604 GAGGACCGGACCATTTCCGCGCGCTGCTACGCGAGGCTGTGCGATGGCGCGCGGTGACC 5663
 Qy 1847 CGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAACTTTGGAACCATCGCCACTTG 1906
 Db 5664 ACCGAGCGGAACCGGTGCGGCTGCCGGGTGACCCGCTGGGTGAAATGCGCTACCTA 5723
 Qy 1907 GCCATTGCAACCCCTCATTTGCGGATGCTTACATGTTGATGGGAAACTAGATTGGGCT 1966
 Db 5724 CCAGTGGCCAGCAGGCGCATTTGCTCCAGCGCCATCTATCCGACGGGTTCACTGCCGCGG 5783
 Qy 1967 GCGATGCCCGTCTTACGCTTGGTCCCAAGATGTGCTTCAAGACCGTGACCTGACCGG 2026
 Db 5784 GCGGCTAAAGCTCCGTCACCTGCGCTGGAATCGTGACGATGGGCTGCGAGCATGTTG 5843
 Qy 2027 CGC---GTGATGTTCTGTGGGCGCAGCGCGGTATCCATTGTCCTCGCGCGGAAGCT 2083
 Db 5844 CGTAGGCTTTCTGTCGCGCCATCACCAGCAGCGACATTTGTCCCGACCAAGAGGCG 5903
 Qy 2084 TTTGTGAGGCAATTGCGCGAGGCTTGGTTGGGACATTTCTCCGAAACCCCAAGCTGT 2143
 Db 5904 TTTACCGCGCAGCGCGCGCGGTGGGATGGGCGCATTTCTCCCGAGAAGCTGGCAGCA 5963
 Qy 2144 CCCATGCTAAAGCAGGAGAGTGTATCTCTCGATGAGATACCCATTGACACCGGATG 2203
 Db 5964 TCTCCGCTTGGCGATGGATTCGTTACGGGTCTCGGACATACACCTCGACGCTCCTC 6023
 Qy 2204 TATTGCAACGATGCGCGCTGGAATCTAGATCTCTAGCTAGACTCAGAGCGCGCTCGTT 2263
 Db 6024 TATTGCAATGCTGGAATCTGACAGTCCGATCGCGCGAATTACCGACACCGGTGAGG 6083
 Qy 2264 GATGCAAGCAATCGAGGATTTGGCGCTTTAGTTTAC 2297
 Db 6084 GCGGCGGCAAGCGCTGTGTACCGGCGCGGCAAC 6117

RESULT 13

AA199683/c

ID AAI99683 standard; DNA; 4403765 BP.

XX*

AC AAI99683;

XX 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX Mycobacterium tuberculosis; strain H37Rv; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence or activity of
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).

XX

SQ Sequence 1200 BP; 272 A; 367 C; 294 G; 267 T; 0 other;

Query Match 4.3%; Score 101; DB 22; Length 1200;
Best Local Similarity 100.0%; Pred. No. 2e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2274 TCGAGGAAATGGCGCCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTGGCCC 2333

Db 1200 TCGAGGAAATGGCGCCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTGGCCC 1141

QY 2334 GCAGGAATGGCGCAGGAGTAACACCTTCAGCAATGG 2374

Db 1140 GCAGGAATGGCGCAGGAGTAACACCTTCAGCAATGG 1100

Search completed: April 27, 2003, 02:57:01
Job time : 10751.4 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 268.11 Seconds
(without alignments)
9634.930 Million cell updates/sec

Title: US-09-105-117K-3

Perfect score: 2374

Sequence: 1 agatactccttggagaaa.....gtaacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/PCT NEW PUB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT NEW PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06 NEW PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06 PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07 NEW PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08 NEW PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08 PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09 NEW PUB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09 PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10 NEW PUB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10 PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60 NEW PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60 PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2374	100.0	3309400	9	US-09-738-626-1
2	993	41.8	993	9	US-09-746-660A-47
3	870	36.6	870	9	US-09-738-626-3456
4	822	34.6	822	9	US-09-746-660A-51
5	708	29.8	708	9	US-09-738-626-3455
6	627	26.4	627	9	US-09-738-626-3454
7	121.2	5.1	909	9	US-09-894-844-12
8	93.2	3.9	988	10	US-09-974-300-77
9	82.6	3.5	894	10	US-09-815-242-6195
10	78	3.3	1077	9	US-09-738-626-1395
11	76.2	3.2	894	10	US-09-815-242-9644
12	73	3.1	903	10	US-09-815-242-4130
13	64	2.7	1277	9	US-09-971-536-15
14	57.6	2.4	762	10	US-09-815-242-7643
15	57	2.4	597	9	US-09-894-844-11
16	43.8	1.8	1973	10	US-09-864-761-3471
17	42.8	1.8	1879	9	US-10-114-170-142
18	40.4	1.7	1075	10	US-09-864-761-19241
19	40.4	1.7	1403	10	US-09-864-761-2513

Sequence 358, App
Sequence 29, Appl
Sequence 1034, Ap
Sequence 358, App
Sequence 20241, A
Sequence 109, App
Sequence 1672, Ap
Sequence 12392, A
Sequence 20, Appl
Sequence 592, App
Sequence 592, App
Sequence 439, App
Sequence 1, Appli
Sequence 346, App
Sequence 346, App
Sequence 7775, Ap
Sequence 34, Appl
Sequence 34, Appl
Sequence 34, Appl
Sequence 1, Appli
Sequence 5, Appli
Sequence 5, Appli
Sequence 398, App
Sequence 188, App
Sequence 398, App

1049 9 US-10-123-155-358
399 10 US-09-861-891-29
431 10 US-09-954-456-1034
1049 9 US-10-123-155-358
1635 10 US-09-864-761-20241
1812 9 US-09-712-363-109
948 9 US-09-738-626-1672
424 10 US-09-960-352-12392
867 9 US-10-123-155-20
499 9 US-10-184-644-592
499 9 US-10-184-644-592
1068 9 US-09-738-626-439
3309400 9 US-09-738-626-1
671 9 US-10-184-644-346
671 9 US-10-184-644-346
996 10 US-09-815-242-7775
4689 9 US-09-860-846-34
4689 9 US-09-988-384B-34
4689 10 US-09-861-289-34
5828 9 US-09-430-029-1
36778 9 US-09-860-846-5
36778 10 US-09-861-289-5
37948 9 US-09-988-384B-5
542 9 US-10-184-644-398
542 9 US-10-123-155-188
542 9 US-10-184-644-398

ALIGNMENTS

RESULT 1

US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 100.0%; Score 2374; DB 9; Length 3309400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGATACTCCTTTGGAGAAACCATGTACCATGTGGTGCATGTTCCTCTGGAAGGC 60
DB 1327595 AGATACTCCTTTGGAGAAACCATGTACCATGTGGTGCATGTTCCTCTGGAAGGC 1327654
QY 61 TCTTTACGTGGGTATTTCCTTACCGTCCAGAGCTCACAGCGGCGGCTGAGTTTAT 120

|||||
Db 1327655 TCCTTACGTGGGATTTCTTCTACGGTCCAGAGCTCACAGCGAGCGCGCTGAGTTCA 1327714
QY 121 GCGGAGAGGGCTGCGCGCTCTTGATTCATCAGCCAGCTATTTCCATCAATTAATCGTTG 180
Db 1327715 GCGGAGAGGGCTGCGCGCTCTTGATTCATCAGCCAGCTATTTCCATCAATTAATCGTTG 1327774
QY 181 GTTGAGAGAACCGGGCGATGACGGTGAGAACTTGTTGCACTGAGCTGCCAACAAATGTCCT 240
Db 1327775 GTTGAGAGAACCGGGCGATGACGGTGAGAACTTGTTGCACTGAGCTGCCAACAAATGTCCT 1327834
QY 241 TGCGCTCATTTGCTTCTCACCCTTGCGAGGGCTGCTCAACGACAAATATCTCGATGG 300
Db 1327835 TGCGCTCATTTGCTTCTCACCCTTGCGAGGGCTGCTCAACGACAAATATCTCGATGG 1327894
QY 301 AATTCCAGAGGGTTCCCGCGCAGCCAGGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360
Db 1327895 AATTCCAGAGGGTTCCCGCGCAGCCAGGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 1327954
QY 361 GAACAAATATTGATATGCTCGCAAGCTCAATGATCATCGCCAGGAACCGCGGCACTCACT 420
Db 1327955 GAACAAATATTGATATGCTCGCAAGCTCAATGATCATCGCCAGGAACCGCGGCACTCACT 1328014
QY 421 TGCGCAGATGGCGCTTGCAATGGGTGCTGCGAGCAAGGAGAGTACGCGCGCGATACCGT 480
Db 1328015 TGCGCAGATGGCGCTTGCAATGGGTGCTGCGAGCAAGGAGAGTACGCGCGCGATACCGT 1328074
QY 481 GACCAATGCAATGATGGTGTCTGTCAGTTGAGCAGCTGGACAAACAGCCCTTGATTCAC 540
Db 1328075 GACCAATGCAATGATGGTGTCTGTCAGTTGAGCAGCTGGACAAACAGCCCTTGATTCAC 1328134
QY 541 CAACAACTTGAGTTTCTGACGCGGAGTTGAGGCGGATCGATGAGATTTCCACAGAGCG 600
Db 1328135 CAACAACTTGAGTTTCTGACGCGGAGTTGAGGCGGATCGATGAGATTTCCACAGAGCG 1328194
QY 601 CGGCATCAACATTTGGGCGAGGCGCACCGATTCCAAACCCGCGGAACTAAACCCATCAA 660
Db 1328195 CGGCATCAACATTTGGGCGAGGCGCACCGATTCCAAACCCGCGGAACTAAACCCATCAA 1328254
QY 661 CATCAGTTGATGCCAATGCGGTTCATCACAACTGCCACAGCAGCTTGATTCAGCGCCA 720
Db 1328255 CATCAGTTGATGCCAATGCGGTTCATCACAACTGCCACAGCAGCTTGATTCAGCGCCA 1328314
QY 721 CACTTTGGGCTGACAGCGGGGTGACAAATGCTGTGCGCGGAAACCCACAGCGGAA 780
Db 1328315 CACTTTGGGCTGACAGCGGGGTGACAAATGCTGTGCGCGGAAACCCACAGCGGAA 1328374
QY 781 CCAGATCAGGCTTGCGCGAAGCGCGCAGCGGGAATCCACCGTCCGGTGTGCGCGTA 840
Db 1328375 CCAGATCAGGCTTGCGCGAAGCGCGCAGCGGGAATCCACCGTCCGGTGTGCGCGTA 1328434
QY 841 TTGCGCCGACGCGCGGATAACACAAACGCGTCAAAATACGCAATTCGGGTTCAACCA 900
Db 1328435 TTGCGCCGACGCGCGGATAACACAAACGCGTCAAAATACGCAATTCGGGTTCAACCA 1328494
QY 901 GGTGACAGCAATGCTCATCAACATGGCTTTACCCAAACCGCTGCTTTATCGAGCTCAC 960
Db 1328495 GGTGACAGCAATGCTCATCAACATGGCTTTACCCAAACCGCTGCTTTATCGAGCTCAC 1328554
QY 961 CTCACCCGACACCGGTTGCGCGTGCAGTGGCCACCGCGCAACCGCCAAAGCGGTGC 1020
Db 1328555 CTCACCCGACACCGGTTGCGCGTGCAGTGGCCACCGCGCAACCGCCAAAGCGGTGC 1328614
QY 1021 ATCGGACAGGTTGTTCTGTTCTCAATGATCTGCGCGCTTCCACTTTGTTGTCAT 1080
Db 1328615 ATCGGACAGGTTGTTCTGTTCTCAATGATCTGCGCGCTTCCACTTTGTTGTCAT 1328674
QY 1081 GGCGTCTTTGCTGCAATGAGCGCAACCAATACAGTATGCGATGCCACCCAGCGCAT 1140
Db 1328675 GGCGTCTTTGCTGCAATGAGCGCAACCAATACAGTATGCGATGCCACCCAGCGCAT 1328734
QY 1141 AATATCGAGCAGATCGGCGCGCATTTGGACAAAGATCAACGCCCAAGGTGCGCGCAT 1200
|||||

Db 1328735 AATATCGAGCAGATCGCGCGCATTTGGACAAAGATCAACGCCCAAGGTGCGCGCAT 1328794
QY 1201 GAACAAAAAGACGTCAAGAAATTAACACACGAGAGAAACCGCAATGAGTCTTTGCGCTT 1260
Db 1328795 GAACAAAAAGACGTCAAGAAATTAACACACGAGAGAAACCGCAATGAGTCTTTGCGCTT 1328854
QY 1261 AATTCCTGTTTAATCAACACAGTACATTTGCGGTCCGATGGACAGTAAAGACGTGCGCCC 1320
Db 1328855 AATTCCTGTTTAATCAACACAGTACATTTGCGGTCCGATGGACAGTAAAGACGTGCGCCC 1328914
QY 1321 CAAAACGACAGCTGAATCAAGATTTCCATGATCACCATGCTGACCTATGGAAGTACTTA 1380
Db 1328915 CAAAACGACAGCTGAATGAAGATTTCCATGATCACCATGCTGACCTATGGAAGTACTTA 1328974
QY 1381 AGTAAATGATTTGTTTCTTAACATGTTTAATATAGCTTTCATGAACCCCATTCAACTGGA 1440
Db 1328975 AGTAAATGATTTGTTTCTTAACATGTTTAATATAGCTTTCATGAACCCCATTCAACTGGA 1329034
QY 1441 CACTTTGCTCTCAATCATGATGAAGGAGCTTCGAGGCGCTCTTAGCCCTTTCCAT 1500
Db 1329035 CACTTTGCTCTCAATCATGATGAAGGAGCTTCGAGGCGCTCTTAGCCCTTTCCAT 1329094
QY 1501 TTCCCTCTCGCGGTGAGTCAAGCGGTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560
Db 1329095 TTCCCTCTCGCGGTGAGTCAAGCGGTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1329154
QY 1561 GGTATCGCGACCCCAACCGGCCAAAGAACCGAAGCGGTGAAGTCTTTGTGCAAGCAGC 1620
Db 1329155 GGTATCGCGACCCCAACCGGCCAAAGAACCGAAGCGGTGAAGTCTTTGTGCAAGCAGC 1329214
QY 1621 GCGGAAATGTTGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTCTGCA 1680
Db 1329215 GCGGAAATGTTGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTCTGCA 1329274
QY 1681 AATCCCGTTAAACCATCGCCATCAACGACATCGCTATCACATGTTTCTCCCGTGT 1740
Db 1329275 AATCCCGTTAAACCATCGCCATCAACGACATCGCTATCACATGTTTCTCCCGTGT 1329334
QY 1741 CAAAGAGGTAGCTTTCTGGGGTGGAGCAACGCTCACGCTTGCGCTTGGAAGATGAAGCGCA 1800
Db 1329335 CAAAGAGGTAGCTTTCTGGGGTGGAGCAACGCTCACGCTTGGAAGATGAAGCGCA 1329394
QY 1801 CACATTTATCTGCTCGCGGTGGAGATGTTTGGAGCGGTAAACCGGTGAGCTAATCC 1860
Db 1329395 CACATTTATCTGCTCGCGGTGGAGATGTTTGGAGCGGTAAACCGGTGAGCTAATCC 1329454
QY 1861 CGTGGCGGATGTGAAGTAGTAGAACTTGGAACCATGCGCCACTTGCGCATTTGCAACCCC 1920
Db 1329455 CGTGGCGGATGTGAAGTAGTAGAACTTGGAACCATGCGCCACTTGCGCATTTGCAACCCC 1329514
QY 1921 CTCATTTGCGGGATGCTTACATGTTGATGGGAACTAGATTTGGGTGCGATGCCGTCTT 1980
Db 1329515 CTCATTTGCGGGATGCTTACATGTTGATGGGAACTAGATTTGGGTGCGATGCCGTCTT 1329574
QY 1981 AGCTTTGCTCCCAAGATGCTTCAAGACCGTGAACCTGGACGCGCGCTGATGGTCC 2040
Db 1329575 AGCTTTGCTCCCAAGATGCTTCAAGACCGTGAACCTGGACGCGCGCTGATGGTCC 1329634
QY 2041 TGTGGGGCGCAGCGCGGTATTCATTTGCTCGCGGGAAGGTTTGGTGAGGCAATTCG 2100
Db 1329635 TGTGGGGCGCAGCGCGGTATTCATTTGCTCGCGGGAAGGTTTGGTGAGGCAATTCG 1329694
QY 2101 CGAGGCTTTGTTGGGACTTCTTCCCGAAACCCAAAGTGTCTCCATGCTAAAGCAGG 2160
Db 1329695 CGAGGCTTTGTTGGGACTTCTTCCCGAAACCCAAAGTGTCTCCATGCTAAAGCAGG 1329754
QY 2161 AGAGTGTCTCTCGATGAGTACCCATTCACACCGATGATTTGGTGGACGATGGCG 2220
Db 1329755 AGAGTGTCTCTCGATGAGTACCCATTCACACCGATGATTTGGTGGACGATGGCG 1329814
QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATTCGAGG 2280
Db 1329815 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATTCGAGG 1329874

;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 3456
;; LENGTH: 870
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3456

Query Match 36.6%; Score 870; DB 9; Length 870;
Best Local Similarity 100.0%; Pred. No. 2.4e-285;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1421 ATCAACCCCAATCACTGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGC 1480
DB 1 ATGAACCCCAATCACTGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGC 60
QY 1481 GCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 1540
DB 61 GCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 120
QY 1541 CATCAGTGGGTGAGTGTGTATCGCGCACCCCAACCGGCCAAAGCAGCGGT 1600
DB 121 CATCAGTGGGTGAGTGTGTATCGCGCACCCCAACCGGCCAAAGCAGCGGT 180
QY 1601 GAAGTCTTGTGAAGCAGCGCGGAAATGGTGTGCTGCAACGACAACTAAAGCGAA 1660
DB 181 GAAGTCTTGTGAAGCAGCGCGGAAATGGTGTGCTGCAACGACAACTAAAGCGAA 240
QY 1661 CTATCTGAGCGCTTGTGTAATCCCGTTAAACCATGCCATCAACGACAGATTCGCTATCC 1720
DB 241 CTATCTGAGCGCTTGTGTAATCCCGTTAAACCATGCCATCAACGACAGATTCGCTATCC 300
QY 1721 ACATGTTTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 1780
DB 301 ACATGTTTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 360
QY 1781 CGCTTGGAGATGAAGCGCACATATCTCTGCTGGCGGTGGAGATGTTTAGGAGCG 1840
DB 361 CGCTTGGAGATGAAGCGCACATATCTCTGCTGGCGGTGGAGATGTTTAGGAGCG 420
QY 1841 GTAAACCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTGAACCTTGAACCATGCGC 1900
DB 421 GTAAACCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTGAACCTTGAACCATGCGC 480
QY 1901 CACTTGGCCATTCGAACCCCTCATTTGGGGATGCCCTACATGTTGATGGGAAACTAGAT 1960
DB 481 CACTTGGCCATTCGAACCCCTCATTTGGGGATGCCCTACATGTTGATGGGAAACTAGAT 540
QY 1961 TGGGCTGCGATGCGGCTTACGCTTCGTTCCCAAGATGTGTTCAAGACCGTGACCTG 2020
DB 541 TGGGCTGCGATGCGGCTTACGCTTCGTTCCCAAGATGTGTTCAAGACCGTGACCTG 600
QY 2021 GACGGGCGCTGATGTCTGTGGGCGCAGCGGTATCCATTGTTCGCTCGGCGGAA 2080
DB 601 GACGGGCGCTGATGTCTGTGGGCGCAGCGGTATCCATTGTTCGCTCGGCGGAA 660
QY 2081 GGTGTTGGTGAAGCAATTCGCGGAGCTTTGGTTGGGAGCTTCTTCCGAAACCCCAAGCT 2140
DB 661 GGTGTTGGTGAAGCAATTCGCGGAGCTTTGGTTGGGAGCTTCTTCCGAAACCCCAAGCT 720
QY 2141 GCTCCCATGTCTAAAGCAGGAGTATCTCTCGATGATATCCCATTTGACACCG 2200
DB 721 GCTCCCATGTCTAAAGCAGGAGTATCTCTCGATGATATCCCATTTGACACCG 780
QY 2201 ATGTATTGGCAACGATGCGCGCTGGATCTAGATCTCTAGCTAGACTCACAGACGCGCTC 2260
DB 781 ATGTATTGGCAACGATGCGCGCTGGATCTAGATCTCTAGCTAGACTCACAGACGCGCTC 840

QY 2261 GTTGATGAGCAATCGAGGATTCGGGCT 2290
DB 841 GTTGATGAGCAATCGAGGATTCGGGCT 870

RESULT 4

US-09-746-660A-51/c
; Sequence 51, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompeius, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zeider, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick
; APPLICANT: Hwang, Byung-Joon
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CP2
; CURRENT APPLICATION NUMBER: US/09/746,660A
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 09/606740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 09/603124
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 51
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(799)
; OTHER INFORMATION: RXA01394
US-09-746-660A-51

Query Match 34.6%; Score 822; DB 9; Length 822;
Best Local Similarity 100.0%; Pred. No. 5.3e-269;
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCAAAACCCGGAATACTAACCCATCAATCAGTTGATGCCCAATCGGTGATC 688
DB 822 GATTCAAAACCCGGAATACTAACCCATCAATCAGTTGATGCCCAATCGGTGATC 763
QY 689 ACAACTGCCACACGACGTTGATCCAGGCCACACCTTGGGGCTGGACAGCGGGGTGAC 748
DB 762 ACAACTGCCACACGACGTTGATCCAGGCCACACCTTGGGGCTGGACAGCGGGGTGAC 703
QY 749 AATGCTGTGCGCCGAAACCCACCGCGGAAACCCAGATCAGGCTTTCGCCGCAACCGCCA 808
DB 702 AATGCTGTGCGCCGAAACCCACCGCGGAAACCCAGATCAGGCTTTCGCCGCAACCGCCA 643
QY 809 GCGGGAATAATCAACGCTCGGTGTGCGCGCCGACGCGCGCGGATTAACACA 868
DB 642 GCGGGAATAATCAACGCTCGGTGTGCGCGCCGACGCGCGCGGATTAACACA 583
QY 869 AACGCGTCAAAATACGATTCGGGTTCAACACGAGGTGACGATTCGATCAACATGGGC 928
DB 582 AACGCGTCAAAATACGATTCGGGTTCAACACGAGGTGACGATTCGATCAACATGGGC 523

```
Qy 929 TTTACCCAAACCCGCTGTTATCGACGCTCACCTCCACCCGACCCGCTTTCGGCGTGTCA 988
Db 522 TTTACCCAAACCCGCTGTTATCGAGGCTCACCTCCACCCGACCCGCTTTCGGCGTGTCA 463
Qy 989 GTGGCCACCGCGCAACCGCCCAAGGGGTGTCTCGGGCAGCGTGGTTCGTCTTCTTCA 1048
Db 462 GTGGCCACCGCGCAACCGCCCAAGGGGTGTCTCGGGCAGCGTGGTTCGTCTTCTTCA 403
Qy 1049 ATGATCTGTGGCGCTTCCACCTGTTGTCTATGCGCTTTCGTGCTGCGCATGACGCGCAAC 1108
Db 402 ATGATCTGTGGCGCTTCCACCTGTTGTCTATGCGCTTTCGTGCTGCGCATGACGCGCAAC 343
Qy 1109 CATAACAGGTAAGCGATGCCCCAGCGCATATATCGAGCAGATCGGCGCGCATTTG 1168
Db 342 CATAACAGGTAAGCGATGCCCCAGCGCATATATCGAGCAGATCGGCGCGCATTTG 283
Qy 1169 GACAAAAGATCAACCGCCAGGTGCGCGCGATGAACAAAAGAGCGTCAGAAATTAACAC 1228
Db 282 GACAAAAGATCAACCGCCAGGTGCGCGCGATGAACAAAAGAGCGTCAGAAATTAACAC 223
Qy 1229 ACAGAAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTGTTTAAATCACCAGTACATTC 1288
Db 222 ACAGAAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTGTTTAAATCACCAGTACATTC 163
Qy 1289 TGGCGTCCGATGGACAGTAAAGACTGGCCCCCAAGAGCAGACCTGTAATGAAGATTTCC 1348
Db 162 TGGCGTCCGATGGACAGTAAAGACTGGCCCCCAAGAGCAGACCTGTAATGAAGATTTCC 103
Qy 1349 ATGATCACCACCGGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAACATGGTT 1408
Db 102 ATGATCACCACCGGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAACATGGTT 43
Qy 1409 TAATATAGTCTTCAAGAACCCCAATCAACTGGACACTTTGCTC 1450
Db 42 TAATATAGTCTTCAAGAACCCCAATCAACTGGACACTTTGCTC 1
```

RESULT 5

```
US-09-738-626-3455/c
; Sequence 3455, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3455
; LENGTH: 708
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3455
```

Query Match 29.8%; Score 708; DB 9; Length 708;
Best Local Similarity 100.0%; Pred. No. 3.4e-230;

```
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 652 ACCATCAACATCAGTTGATGCGCAATGCGGTTCATCAAACTGCGACGACGTTGAT 711
Db 708 ACCATCAACATCAGTTGATGCGCAATGCGGTTCATCAAACTGCGACGACGTTGAT 649
Qy 712 CAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGAACAATGCTGCTGCGCCGAAACCCAC 771
Db 648 CAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGAACAATGCTGCTGCGCCGAAACCCAC 589
Qy 772 CAGCGGAACACAGATCAGGCTTGC CGGAAACGCGCCAGCGGGGAAATCAACGCTCCGT 831
Db 588 CAGCGGAACACAGATCAGGCTTGC CGGAAACGCGCCAGCGGGGAAATCAACGCTCCGT 529
Qy 832 GTCGCGTATGCGCGCGACGCGCGGATGAACAACAAACGCGTCCAAATACGATTCCG 891
Db 528 GTCGCGTATGCGCGCGACGCGCGGATGAACAACAAACGCGTCCAAATACGATTCCG 469
Qy 892 GTTCAACACAGGTGAGCAGCATGATGCGCATCAACATGGGCTTTACCCAAACCGCTGCTTATC 951
Db 468 GTTCAACACAGGTGAGCAGCATGATGCGCATCAACATGGGCTTTACCCAAACCGCTGCTTATC 409
Qy 952 GAGCTCACTCCACCGCACCGGTTGCGGTGTCAGTGGGCGACCGCGAAACCGCCAA 1011
Db 408 GAGCTCACTCCACCGCACCGGTTGCGGTGTCAGTGGGCGACCGCGAAACCGCCAA 349
Qy 1012 AGCGGTGTCATCGGCGACGCGTGGTTCGTTCTTCAATGATCTGCGCTTCCACCTT 1071
Db 348 AGCGGTGTCATCGGCGACGCGTGGTTCGTTCTTCAATGATCTGCGCTTCCACCTT 289
Qy 1072 GTTGTGTCATGGGCTCTTTGCGTCGATGAGCGCAACCAATACAGTAAGCGATGCCACC 1131
Db 288 GTTGTGTCATGGGCTCTTTGCGTCGATGAGCGCAACCAATACAGTAAGCGATGCCACC 229
Qy 1132 CMAGCGCAATATCGAGCAGCATGCGCGCGCATTTGGACAAAGATCAACGCCCAAGGT 1191
Db 228 CMAGCGCAATATCGAGCAGCATGCGCGCGCATTTGGACAAAGATCAACGCCCAAGGT 169
Qy 1192 GCGGCGATGAACAAAGACGTCAGAAATTAACACACAGAGNACCGCAATGATGCC 1251
Db 168 GCGGCGATGAACAAAGACGTCAGAAATTAACACACAGAGNACCGCAATGATGCC 109
Qy 1252 TTCGCGCTTAATTCCTTGTGTTAATCAGCATGATCTTCTCGGTCCGATGAGCAAGAG 1311
Db 108 TTCGCGCTTAATTCCTTGTGTTAATCAGCATGATCTTCTCGGTCCGATGAGCAAGAG 49
Qy 1312 ACTGGCCCCCAAGACGACCTGTAATGAAGATTTCCATGATCACCAT 1359
Db 48 ACTGGCCCCCAAGACGACCTGTAATGAAGATTTCCATGATCACCAT 1
RESULT 6
US-09-738-626-3454
; Sequence 3454, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3454
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3454

Query Match 26.4%; Score 627; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 1.2e-202; Mismatches 0; Indels 0; Gaps 0;
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 ATGTACGCAATGGGTGACATTTGGTCTCGAAGAGCTCTTACGTGGGTATTTCTTC 82
DB 1 ATGTACGCAATGGGTGACATTTGGTCTCGAAGAGCTCTTACGTGGGTATTTCTTC 60
QY 83 TAGGTTCCAGAGCTCACAGCGAGGCGGCTGAGTTATGCGGAGGAGGCGTCCCGCTT 142
DB 61 TAGGTTCCAGAGCTCACAGCGAGGCGGCTGAGTTATGCGGAGGAGGCGTCCCGCTT 120
QY 143 CTGATTCATCAGCAAGCTATTCATCAATTAATGTTGGTGGAGGAACCGGCGATGAC 202
DB 121 CTGATTCATCAGCAAGCTATTCATCAATTAATGTTGGTGGAGGAACCGGCGATGAC 180
QY 203 GGTGAGAACTTTGTGAGTCAAGTCCCAACAAATGCTTGGCGTCAATTTCTCACCA 262
DB 181 GGTGAGAACTTTGTGAGTCAAGTCCCAACAAATGCTTGGCGTCAATTTCTCACCA 240
QY 263 CTTGCGCAGGCGCTGCTCAGGCAAAATATCTGATGGAATTCAGAGGTTCCGCGGC 322
DB 241 CTTGCGCAGGCGCTGCTCAGGCAAAATATCTGATGGAATTCAGAGGTTCCGCGGC 300
QY 323 AGCCAGGTAAGTCCCTGCTGAGGCGATGTTGAACGTGAACAAATATGATATGTCGCG 382
DB 301 AGCCAGGTAAGTCCCTGCTGAGGCGATGTTGAACGTGAACAAATATGATATGTCGCG 360
QY 383 AAGTCAATGACATCGCCAGGAACCGGCGAGTCACTTCGCGAGATGGCGTTGCATGG 442
DB 361 AAGTCAATGACATCGCCAGGAACCGGCGAGTCACTTCGCGAGATGGCGTTGCATGG 420
QY 443 GTGTCGCGCAGCAGAGAGTACGCGCGGATACCGTACCGAGTGAATGATGTTGCT 502
DB 421 GTGTCGCGCAGCAGAGAGTACGCGCGGATACCGTACCGAGTGAATGATGTTGCT 480
QY 503 TCCTCAGTTGAGCAGCTGGACACAGCCTTGATTCTACTCAACACTTGGAGTTTCTGAC 562
DB 481 TCCTCAGTTGAGCAGCTGGACACAGCCTTGATTCTACTCAACACTTGGAGTTTCTGAC 540
QY 563 GCCGAGTTGGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGCGAAG 622
DB 541 GCCGAGTTGGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGCGAAG 600
QY 623 GCCACCGATTCAAAACCCCGGAAAAC 649
DB 601 GCCACCGATTCAAAACCCCGGAAAAC 627

RESULT 7

US-09-894-844-12
; Sequence 12, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844

; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 909
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-09-894-844-12

Query Match 5.1%; Score 121.2; DB 9; Length 909;
Best Local Similarity 49.5%; Pred. No. 3.8e-30;
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;
QY 1433 CAACTGGACACATTTCTCTCAATCATTTGATGAAGCAGCTTCGAAGGCGCTCTTAGCC 1492
DB 28 CAGCTGGCGCATTTGGTCCGCTGCTGAACTTGGGCGAGCTTCGATCGGCGCGGAGCGC 87
QY 1493 CTTTCCATTTTCCCTCGGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTGGGT 1552
DB 88 CTACATGTCACCCCGTCTGCTCAGTCAGCGCATCAAGTCGTTGGAGCAGCAGGTCCGC 147
QY 1553 CGAGTGTGGTATCGGCAACCCCAACCGGCCAAAGCAACGGAAGCGGCTGAATCTTTGTG 1612
DB 148 CAGGTGCTGTGGTTCAGGGAAGGCAATGTCGGGCGGCGACCGCAGGTATCCCGCTGTTG 207
QY 1613 CAAGCAGCGCGAAATGTTGTTGCTGCAAGCAGAAACTTAAAGCGCAACTATCTCG--AC 1670
DB 208 CGTGGCGCGCGCAACAGCGTTGCTCGAGTCCGAGGCGCTCGCTGAATGAGTGGCAAC 267
QY 1671 GCGTTGCTGAA----ATCCCGTTAAACCATTCGCAATCAACGCAAGATTCGCTATCCACATGG 1726
DB 268 GCGTCGTGAAACGCGCGGATCACCATTGCGTTAAACCGCATTCATCGCGCATGG 327
QY 1727 TTTCTCCCGCTGTTCAACGAGGTAGCTTTCTGGGGTGGAGCAACGCTACGCTGCGCTTG 1786
DB 328 TTTCTCCCGCTGTTGAGC-----GTCTCGGCGACGCTCTGCTGACGTTCCGATC 378
QY 1787 GAAGTGAAGCGCACACATTTATCTTCTGCGCGCTGGAGATGTTTATAGAGCGGTAAAC 1846
DB 379 GAGGACCGAACAATTCGCGCGCTGCTACGCGAGGCTGCGCATGCGCGCGGTGACC 438
QY 1847 CGTGAAGCTTAATCCCGTGGCGGATGTAAGTAGTAGAATTTGGAACCACTTGCCCACTTG 1906
DB 439 ACCGAGCGGAACCCCGTGCCTGCGGCTGCGCGCTGCGCTGGAATGCGCTACCTA 498
QY 1907 GCAATTTGCAACCCCTCATTTGCGGGATGCTTACATGTTGATGGGAACCTAGATTGGGCT 1966
DB 499 CCAGTGGCGCAGCAGGCGCAATTCGTCAGCGCCATCTATCCGAGCGGTTCACTGCCCGCG 558
QY 1967 GCGATGCCCGCTTACGCTTCCGCTCCCAAGATGCTTTCAAGACCGTGCACCTGACGCG 2026
DB 559 GCGGCTAAAGCTCCGTCATCTGGGTGGAATCGTAGATGGGCTGCGAGCATGTTGGTG 618
QY 2027 CGC---GTCGATGCTCTGTGGGCGCAGCGCGCTATTCATTTGTCGCGCGGAAGGT 2083
DB 619 CGTAAGGCTTTCTGTCGCGCATCACCACGCGCACTTTTGTCCCGACCAAGAGGCG 678
QY 2084 TTTGTTGAGGCAATTCGCGAGCGCTTGGTTGGGAGCTTCTTCCCGAAACCCAGCTGCT 2143
DB 679 TTTACCGCGCGCAGCGCGCGCTGGGATGGGCGATGTTCCCGGAGAGCTGGCAGCA 738
QY 2144 CCAATGCTAAAGCAGGAGAGTGAATCTCTCGATGATATCCCATTCGACACCGCATG 2203
DB 739 TCTCCGCTTGGCGATGATGCTGTTGACGGGTCTGGGACATACCTCGAGCTCCCTCTC 798
QY 2204 TATTTGCAACGATGGCGCTCGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTT 2263
DB 799 TATTGGCAATGCTGGAACATGGAACAGTCGATCATCGCGCAATTAACCGACACGCGTGAG 858

QY 2264 CATGAGCAATCGAGGGATTCCGGCTTAGTTAC 2297
 Db 859 CGCGCGCAAGCGTCTGTACCGGGCCAGCAAC 892

RESULT 8

US-09-974-300-77
 ; Sequence 77, Application US/09974300
 ; Patent No. US20020146721A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Berka, Randy M.
 ; APPLICANT: Clausen, Ib Groth
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 ; FILE OF INVENTION: Expression
 ; FILE REFERENCE: 10085.500-US
 ; CURRENT APPLICATION NUMBER: US/09/974.300
 ; CURRENT FILING DATE: 2001-10-05
 ; PRIOR APPLICATION NUMBER: 09/680,598
 ; PRIOR FILING DATE: 2000-10-06
 ; PRIOR APPLICATION NUMBER: 60/279,526
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 8481
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 77
 ; LENGTH: 988
 ; TYPE: DNA
 ; ORGANISM: Bacillus licheniformis
 US-09-974-300-77

Query Match 3.9%; Score 93.2; DB 10; Length 988;
 Best Local Similarity 56.0%; Pred. No. 1.4e-20;
 Matches 251; Conservative 0; Mismatches 173; Indels 24; Gaps 3;
 QY 3 ATACTCTCTTGGAGAACCATGATGACGATTCGTCGATGTTGGCTCTGGAAGCTC 62
 Db 428 ATACCGCTGGAAGAGACGATGGAGCGCTCGCCCATCGTCAGGCAAGAAAGCTT 487
 QY 63 TTTCAGTGGGATTTCCTTACCGTTCAGAGCTTCACAGCGAGCGGCTGAGTTCATGG 122
 Db 488 TATATGTCGGTTTGTCTGAACCTACAGTGTGAACAGACCGAAAGCGGCTGATTTTC 547
 QY 123 CGGAGGAGGCTGCGGCTTCTGATTCATCAGCAAGCTATTCATCATTAATGTTGGG 182
 Db 548 AAGAATACGGGATTCGCTCTCTCATCCATCAGCCCAATATTCATGTTCCACAG----- 602
 QY 183 TGGAGGAACCGGGGATGACGGTGAAGACTTGTTCAGTCTGAGTCCCAACAATGCTTGT 242
 Db 603 -----AGAACCGGAAAGGCTGCTGCAAGTGTGCGAG-----AAAAAGGAGCCG 649
 QY 243 GCGTCATTGCTTTCTACCACTTCGCGAGGCGCTGCTCAGGACAAATATCTCGATGAA 302
 Db 650 GGGGATCGGCTTTACACCTCTTCCCAAGGGCTTTTAAACAAGAAATATTTGACAGGA 709
 QY 303 TTCAGAGGTTCCGCGCCAGCCAGG-----GTAAAGTCCCTGCTGAGGGCATGTGA 356
 Db 710 TTCCCGCGGGTCAAGAGCCGCTGATCGAACAGTCCGTTTTTCAGGCGCCGAGGCATCA 769
 QY 357 ACGTGAACATATTGATGTGTCGCAAGCTCATGATGATCGCCACAGGAGCGGGCAGT 416
 Db 770 CAGAAGACGTGCTCGGCAAGTTCGCAAGCTGAATCAAGTGGTGTGCCGAAACGCGGCCAGT 829
 QY 417 CACTTGGCGAGATGGCGCTTGCATGGGT 444
 Db 830 CTCTTCGCAATGCGATAGCTGGCT 857

RESULT 9

US-09-815-242-6195
 ; Sequence 6195, Application US/09815242
 ; Patent No. US20020061569A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815.242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6195
 ; LENGTH: 894
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(894)
 US-09-815-242-6195

Query Match 3.5%; Score 82.6; DB 10; Length 894;
 Best Local Similarity 48.8%; Pred. No. 5.3e-17;
 Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;
 QY 1455 TCATTGATGAAGGACGCTTCGAAGGCGCTCTCTAGCCCTTTCCATTTCCCTCGGCG 1514
 Db 44 TGATACGTGAACGAGGATTTGAGCGCGCGCACAAAGCTGTGCATTACACAATCAGCG 103
 QY 1515 TGAGTCAGCGCGTTAAAGCTCTCGACATCATCGTGGTCAAGTGTGGTATCGGCACCC 1574
 Db 104 TCTCAGACGCGCATTAAGCAACTTGAATAATATGTTCCGGCAGCGCTGTTGGTCCGTACCG 163
 QY 1575 AACCGGCCAAAGCAACGAGCGGGTGAAGTCTTGTGCAAGCAGCGCGGAAATGGTGT 1634
 Db 164 TACCGCGCGCGGACGGAACGAGGCAAAACTGCTGGCACTGCTGCCAGGTGGAGT 223
 QY 1635 TGCTGCAAGCAGAAACTAAAGCGCAACTATCTGAGCGCTTCTGAAATCCCGTTAACCA 1694
 Db 224 TGCTGGAAGAGAGTGGCTGGCGATGAACAAACCGTTTCGACTCCGCTGCTGCTTAC 283
 QY 1695 TCGCCATCAACGAGATTCGCTATCCATATGTTTCTCCCGTGTTCACAGGAGTAGTT 1754
 Db 284 TGGCGGTCAACGCCGACAGTCTGGCGACGTGTTGTTCTCTGCACTGGCTCTCTGTTGG 343
 QY 1755 CTTGGGGTGGAGCAACGCTCACGCTCGCTTGAAGATGAAGCGCACACATTCCTTGC 1814
 Db 344 CTGATTGCGCTATCCGCTCAACTTGCAGGTAGAAGATGAACCCGCACTCAGGAACGTC 403
 QY 1815 TGGCGGTGGAGATGTTTGGAGCGGTAAACCCGTGAAGCTAATCCCGTGGCGGATGTG 1874
 Db 404 TGGCGCGCGCGAAGTGGTGGCGGCTGAGTATTCAACATCAGGGCTGCCGAGTTGTC 463
 QY 1875 AAGTAGTAAGTCTGGAACCATCGGCCACTTGGCCAT 1911
 Db 464 TTGTCGATAAAGTCTGGTGGCTCGACTATCTGTTGT 500

RESULT 10

US-09-738-626-1395/c
; Sequence 1395, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIRO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 1395
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: *Corynebacterium glutamicum*
US-09-738-626-1395

Query Match 3.3%; Score 78; DB 9; Length 1077;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2297 CTCTGAAAGGTTTCAGGTTTTCATCTTCCCGCGAGGAATGGCGCAGGAGT 2356
DB 1077 CTCTGAAAGGTTTCAGGTTTTCATCTTCCCGCGAGGAATGGCGCAGGAGT 1018
QY 2357 AACACCTTCAGCAATGG 2374
DB 1017 AACACCTTCAGCAATGG 1000

RESULT 11

US-09-815-242-9644
; Sequence 9644, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITFA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 9644
; LENGTH: 894
; TYPE: DNA
; ORGANISM: *Salmonella typhi*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(894)
US-09-815-242-9644

Query Match 3.2%; Score 76.2; DB 10; Length 894;
Best Local Similarity 44.8%; Pred. No. 8e-15;
Matches 340; Conservative 0; Mismatches 413; Indels 6; Gaps 1;

QY 1423 GAAACCCATTCAACTGGACACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAAGGCGC 1482
DB 12 CGACTACAGAACACTACAGGCGCTGATCGCTTATACGTGAACGAGGATTCGAGCGCGC 71
QY 1483 CTCTTAGCCCTTTCCATTTCCCTCGGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCA 1542
DB 72 CGCCCCAAGAGCTGTGCATCAGCAATCCGCGCTCTCACAGCGCATATAAAACAGCTTGAGAA 131
QY 1543 TCACGTGGGTGAGTGTGTTGATCGCGCACCAACCGGCGCAAGCAACGAAAGCGGTCGA 1602
DB 132 TATGTTGAGCAACACTTTTGTGTCGTACCGTTCTCTCGCCCCACGAGCAAGGGCA 191
QY 1603 AGTCCTTGTGCAAGCAGCGCGGAAATGTTGTGTGCAAGCAGAACTATAAGCGCACT 1662
DB 192 AAAATTTACTGGCTCTGCTGCGTCAGCTTGAACCTGCTGGAAGAGAGTGCTGGCGATGA 251
QY 1663 ATCTGAGCGCTTGTGAATCCGTTAACCATCGGCATCAAGCGCAGATTCGCTATCCAC 1722
DB 252 ACAACCGGCTCTACGCGCTACTGCTGCTGCGGCTTAAACGCGCAGACTCTGCGCAC 311
QY 1723 ATGGTTTCTCCGCTGTTTCAACGAGGTAGCTTTTGGGGTGGAGCAACGCTCACGCTGG 1782
DB 312 CTGGCTGCTTCGGCGCTGGCGCCAGTACTGGGGAATCCCTATCCGCTCTCAATTTACA 371
QY 1783 CTGGAAGATGAAGCGCACACATATCTTGTGCGGCGGTGAGATGTTTATAGGAGCGGT 1842
DB 372 GGTGGAAGATGAACGCGCACAGACGCTCTGCGTCGCGCGAGGTGGTGGCGCGGT 431
QY 1843 AACCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAATTTGGAAACCATGGCGCA 1902
DB 432 GAGTATTCAGCATCAGGCGCTGCCAGTTGCTGTGTAGACAAACTGGGCGCGCTCGACTA 491
QY 1903 CTGGCCATTGCAACCCCTCATTTGGGATGCTTACATGTTGATGGGAAACTAGATTG 1962
DB 492 TCTGTTTGTGCGGTCAAGCCGTTTTCGCGAGCGCTATTTTCCCAATGGGTTACCCGGTC 551
QY 1963 GGTGCGAT-----GCCCGTCTTAGCTTCGTTCCGTCCTCAAGATGTCTTCAAGACCGTA 2016
DB 552 GTCGCTGTGAAAGCGCGCGCTGCTGCTGCGTTTCGACCATCTGGATGATATGCATCAGGCTTT 611
QY 2017 CTGGAGCGCGGTGATGTTGCTGTGGGCGCGAGGCGGTATCCATTTGTCCTCGCGC 2076
DB 612 TTGCAACAGAAATTTGCATCTACCGCGCGCAGCGCTGCGCTATCGTTAAATTCATC 671
QY 2077 GGAAGGTTTTGTGAGGCAATTCGCGAGGCTTGTGTTGGGACTTCTTCCCGAAACCCA 2136
DB 672 CGAAGGTTTTGTACAGCTTGGCGGTGAGGCGAGCGCTGCTGATGATCCCTCACCTGCA 731
QY 2137 AGCTGCTCCCATGCTATAAAGCAGGAGAGTATCTCTCT 2175
DB 732 GATTGAGAAAGAGCTGAAAGCGCGGAGTTGATCAACTT 770

RESULT 12

US-09-815-242-4130
; Sequence 4130, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4130
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4130

Query Match 3.1%; Score 73; DB 10; Length 903;
Best Local Similarity 46.0%; Pred. No. 9.9e-14;
Matches 413; Conservative 0; Mismatches 460; Indels 24; Gaps 4;

QY 1418 TTCAATGAACCCCACTTCAACTGGACACTTGTCTCTCAATCATTTGATGAAGCAGCTTCGAA 1477
DB 7 TTTGTCTACTACAAAGTTGCTCGCCGCGTGGCGCGGTGGAGAGGCGGTTTCGAG 66
QY 1478 GCGCGCTCTTAGCCCTTTCCATTTCCCTCGCGCGGTGAGTCAGCGCGTTAAAGCTCTC 1537
DB 67 CGCGCCGACACAGGCGCTGGGCTGTGCAATCGCGGTATCGCAGCGGATCAAGCTGCTC 126
QY 1538 GAGCATCAGTGGTTCAGTGTGGTATCGCGCACCCAAACCGGCAAGCAACCGAAGCG 1597
DB 127 GAGCGCGGCTCGGCCACAGCGGCTCTGTGGCGGAGACGCGCGCATCCCAACCGATCTC 186
QY 1598 GGTGAAGTCTTTGCAAGCAGCGCGAATAATGTGTGTCGCAAGCAGAACTAAG-- 1655
DB 187 GGGCGCCCTGTCTCAACCATGTGACAGGTGCGGTGCTGGAGGCGACCTTCGACGCGC 246
QY 1656 -----CGCAACTATCTGGACGCTTGTCTGAATCCGTTAAACCATCGCCATCAACGCA 1708
DB 247 TGGTGCAGAACTTGGATGAAGGGGTGCCCCGGAACGCTGCGCATCGTCTCAACGCC 306
QY 1709 GATTCGCTATCCATGTTCTTCCCGTGTTCACGAGTAGTCTTCTGGGGTGGAGCA 1768
DB 307 GATAGCTGGCGCACTGTGGCGCGCGGTGGCGGATTTCTGCGCGCAGCGCGGGTG 366
QY 1769 ACGTCAAGTGGCTGGGAAGATGAGCGCACATATTCTCTGTCGCGGTGGAGAT 1828
DB 367 CTGCTGAGACCTTGTGGTGGAGGACCAAGAGGTAGGGCTCAAGCGCATGCGCGGGCGAA 426
QY 1829 GTTTTAGAGCGGTAAACCCGTGAAGCTTAATCCCGTGGCGGATGTGAAGTAGTAGAATT 1888

DB 427 GTGCGCGGCTGCTGCTCGGTAGCGCGCGCGGTGCGCGCGCGGTGCTGCTGCTC 486
QY 1889 GGAACCATGCGCCACTTGGCCATTTGCAACCCCTCAAT-----GCGGATGCTTACATG 1942
DB 487 GGAGCCATGCGTTATCGCGGCTTGGCCAGCCCGGATTTCAATTGCCCGGCACTTTCCCGCG 546
QY 1943 GTTGATGGGAAACTAGATTGGGCTGCGATGCGCGCTTACGCTTCCGTTCCCAAGATG 2002
DB 547 GGGTCTGAGCGCGCCCTTGGCGCGGTGCGCGCATCGTGTTCGGCCCGGACGACCTG 606
QY 2003 CTTCAAGACCGGTGACCTGGACGCGCGCTCGATGTCCTGTGCGGCGCAGGCGCGTATCC 2062
DB 607 CTGCAACACCGCTTCTCAAGGACCTCGCG-----TCGAGGCGGCTTTCATCCACAC 660
QY 2063 ATTGTCCGTCGCGGGAAGGTTTGTGAGGCAATTCGCGGAGGCTTGTGGTGGGACTT 2122
DB 661 CTCTGCGCGTCTCGGAAGGTTTCTGTCGCTCACCGCGCGGCTCTCGGCTGGGCGCTG 720
QY 2123 CTTCCCGAAACCCCAAGCTGCTCCCATGCTAAAGCAGGAGAGTATGTC---CTCCTCGAT 2179
DB 721 GTGCCGAGCGCGGCTGCAAGCGAGCTGGCGCGCGGAACTGGTCAAGCTGCTGCGCC 780
QY 2180 GAGATACCCATTGACACACCGATGTTATGGCAACCATGCGCTGGAATCTAGATCTCTA 2239
DB 781 GGCCAGGTCTACGAGCTACCGTTGTTACTGGCACTACTGCGGCAACGCGCGGCACTGCTC 840
QY 2240 GCTAGACTCACAGACGCGCTGTTGATGAGCAATCAGGGATTGCGGCTTAGTTA 2296
DB 841 GCGTCTGCTCACCGAGCACTGCTGCGCGCGCGGCTTGGTGGGGTGCTCA 897

RESULT 13

US-09-971-536-15
; Sequence 15, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:

; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; FILE REFERENCE: 1043C2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1277
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-15

Query Match 2.7%; Score 64; DB 9; Length 1277;
Best Local Similarity 47.9%; Pred. No. 1.5e-10;
Matches 279; Conservative 0; Mismatches 270; Indels 33; Gaps 2;
QY 9 CTTTGAAGAAACCATGTACGCAATTCGCTGACATTTGCGTCTCGAAAGGCTCTTTACG 68
DB 608 CTTTGAAGAGAGCGTTAATGCGCTGGATCAACCGTGGTGACGTTAGGCTGTGTTATA 667

GenCore version 5.1.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:22:28 ; Search time 133.755 seconds
(without alignments)
5443.169 Million cell updates/sec

Title: US-09-105-117K-3
Perfect score: 2374
Sequence: 1 agatactcttgggaagaa.....gtacaccttcagcaatgg 2374

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/6C_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122.8	5.2	15239	1	US-08-390-878-17
2	121.2	5.1	4403765	4	US-09-103-840A-2
3	121.2	5.1	4411529	4	US-09-103-840A-1
4	100.8	4.2	5541	1	US-08-920-812-20
5	100.8	4.2	5541	1	US-08-920-827-20
6	100.8	4.2	5541	1	US-08-921-177-20
7	100.8	4.2	5541	1	US-08-362-577C-20
8	100.8	4.2	5541	2	US-08-920-828-20
9	42.8	1.8	1879	4	US-09-453-702B-142
10	39.6	1.7	1935	2	US-08-492-027A-9
11	37.6	1.6	4403765	4	US-09-103-840A-2
12	37.6	1.6	4411529	4	US-09-103-840A-1
13	37	1.6	2058	1	US-08-358-117-1
14	37	1.6	2058	3	US-08-470-588-1
15	35.4	1.5	1149	3	US-09-134-001C-928
16	35	1.5	4689	4	US-09-105-537-34
17	35	1.5	36778	4	US-09-105-537-5
18	35	1.5	38506	3	US-09-320-878-19
19	34.6	1.5	6611	1	US-08-402-282-2
20	34.6	1.5	6611	1	US-08-508-004-2
21	34.6	1.5	6611	1	US-08-402-066-2
22	34.6	1.5	6611	1	US-08-402-068-2
23	34.4	1.4	2048	3	US-08-776-246-1
24	34.4	1.4	2048	3	US-08-776-251-1
25	33.6	1.4	1881	4	US-09-434-288-7
26	33.6	1.4	8438	4	US-07-945-283-1
27	33.2	1.4	6133	4	US-09-453-702B-15

28	33.2	1.4	20235	1	US-07-642-734C-3	Sequence 3, Appli
29	33.2	1.4	20235	3	US-08-439-009A-3	Sequence 3, Appli
30	33	1.4	379	4	US-08-205-697A-27	Sequence 27, Appl
31	33	1.4	379	4	US-08-702-525-27	Sequence 27, Appl
32	33	1.4	379	5	PCT-US95-02576-27	Sequence 27, Appl
33	33	1.4	707	4	US-08-998-416-940	Sequence 940, App
34	33	1.4	1491	2	US-08-147-772-1	Sequence 1, Appli
35	33	1.4	1491	2	US-08-456-104-5	Sequence 5, Appli
36	33	1.4	1491	2	US-08-101-624-22	Sequence 22, Appl
37	33	1.4	1491	2	US-08-751-767A-5	Sequence 5, Appli
38	33	1.4	1491	3	US-08-153-262-1	Sequence 1, Appli
39	33	1.4	1491	3	US-08-479-744A-28	Sequence 28, Appl
40	33	1.4	1491	3	US-08-280-757B-28	Sequence 28, Appl
41	33	1.4	1491	3	US-09-159-135-1	Sequence 1, Appli
42	33	1.4	1491	4	US-08-205-697A-18	Sequence 18, Appl
43	33	1.4	1491	4	US-08-702-525-18	Sequence 18, Appl
44	33	1.4	1491	4	US-09-450-798-1	Sequence 1, Appli
45	33	1.4	1491	4	US-09-326-186B-225	Sequence 225, App

ALIGNMENTS

RESULT 1
US-08-390-878-17
; Sequence 17, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-17

Query Match 5.2%; Score 122.8; DB 1; Length 15239;
Best Local Similarity 49.7%; Pred. No. 9e-26;
Matches 434; Conservative 0; Mismatches 422; Indels 18; Gaps 4;

QY 1433 CAACTGGACATTCTCTCAATCATTTGATGAAGCAGCTTCGAAAGGCGCTCTTACCC 1492
DB 5253 CAGCTGGCGGCGATTGGCTGCTCGCGTGTTCGAACTGGCGAGCTTCGATCGCGCGCGAGGCG 5312

```
QY 1493 CTTTCCATTTCCTCCCGGTCAGTCAGCGGTTAAAGCTCTCGAGCATCAGTGGGT 1552
Db 5313 CTACATGTCACCCCTCCGCTGTGAGTCAGCGCATCAAGTCGTTGGAGCAGCGTCGGC 5372
QY 1553 CGAGTGTGGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGCTGAAGTCTTGTG 1612
Db 5373 CAGGTGTGGTGTGTCAGGGAAGCCATGTGGGCGAGCGCAGGATGATCCGCTGTG 5432
QY 1613 CAAGCAGCGGGAATAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGG--AC 1670
Db 5433 CGGTGGCGCGCAACAGAGGTTGCTCGAGTCCGAGCGCTGCTGAAATGGGTGGCAAC 5492
QY 1671 GCCTTGTCTGAA---ATCCGCTTAACCATCGCATCAACGAGATTCGCTATCCATGG 1726
Db 5493 CGGTGCTGAAACCGCACCGCGATCACCATTGGGTAACCGCGGATTCATGGGACATGG 5552
QY 1727 TTTCTCTCCGCTGTTCAACGAGGTAGCTTTCTGGGGTGGAGCAACGCTCACGCTGCGCTTG 1786
Db 5553 TTTTCGCGCGTGTTCGAGC-----GTCTCGCGCACTGCTCTCGAGCTTCGATC 5603
QY 1787 GAAGATGAAGGCGACATATTCCTGCTGGGCGTGGAGATGTTTGGAGCGGTAAAC 1846
Db 5604 GAGGACGAGACCATTCCTCGCGCGCTCTACCGGAGGTTGTGCGATGGGCGCGGTGACC 5663
QY 1847 CGTGAAGCTAATCCGCTGGCGGATGTGAAGTAGTAGAATCTTGGAACTTGGAACTTGG 1906
Db 5664 ACCGAGCGGAACCGGTGCGCGGCTGCGGGGTGACCCGCTGGGTGAAATGCGCTACCTA 5723
QY 1907 GCCATTGCAACCCCTCATTTGCGGATGCTTACATGTTGATGGGAACTAGATTGGGCT 1966
Db 5724 CCAGTGGCCAGCAGGCGCATTCGTCAGCGCATCTATCCGACGGTTCACTGCGCGCGG 5783
QY 1967 GCGATGCCGCTTACGCTTCGCTGCCGCGGCTGGGATGGGCGATGTTCCCGAGCAGAG 5903
Db 5784 GCGGCTAAAGCTCCGCTCACTGGCGGTGAATCGTGACATGGGCTGCGAGCATGTTGGTG 5843
QY 2027 CGC---GTCCGATGCTGTGGGCGCGAGCGGTATCCATTCTCCGTCGGCGGAAGT 2083
Db 5844 CGTAAGCGCTTGTGTCGGCATACACAGACCGACGCTTGTCCCGACCAAGAGGCG 5903
QY 2084 TTTGGTGAAGCAATTCGCGAGGCTTGGTTGGGAGCTTCTTCCCGAAACCCAAAGCTGCT 2143
Db 5904 TTCACCGCGCAGCGCGCGGCTGGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 5963
QY 2144 CCATGCTAAAGCAGAGAGATGATCTCTCGATGAGATACCCATTTGACACCGCATG 2203
Db 5964 TCTCCGCTTCCGATGATCGTTCTGATCGGCTCTGCGACATACACCTCGACGCTCCCTC 6023
QY 2204 TATTGGCAACGATGGCGCTGGAACTAGATCTCTAGCTAGACTCACAGACCGGCTGTT 2263
Db 6024 TATTGGCAATGCTGGAACCTGGACATCCGATCCGATCATCGCGCAATTAACGACCGT 6083
QY 2264 GATGAGCAATCGAGGATTCGCGCTTAGTTAC 2297
Db 6084 GCGGCGCAAGCGCTGTGTACCGGCGCGCAAC 6117
```

RESULT 2

```
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match          5.1%; Score 121.2; DB 4; Length 4403765;
Best Local Similarity 49.5%; Pred. No. 7.1e-24;
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;

QY 1433 CAACTGGACATTTGCTCTCAATCAATTGATGAAGCAGCTTCGAAGGCGCTCTCTTAGCC 1492
Db 2227172 CAGCTGGCGCATTTGGCTGCGCTGCTGCAACTGGGCGAGTTCGATGCGCGCGGAGCGC 2227113
QY 1493 CTTTCCATTTCCCTCCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTGGGT 1552
Db 2227112 CTACATGTCACCCCGTCCGCTGTGAGTCAGCGCATCAAGTCGTTGGAGCAGCAGGTCCGC 2227053
QY 1553 CGAGTGTGGTATCGCGCACCCCAACCGGCGCAAGCAACCGAAGCGGCTGAAGTCTCTTGTG 1612
Db 2227052 CAGGTGCTGGTGGTTCAGGGAAGCCATGTCGGGCGAGCAGCGAGGTATCCCGCTGTTG 2226993
QY 1613 CAAAGCAGCGCGGAAATGTTGTTGTCGAAGCAGAAACTAAAGCGCAACTATCTGG--AC 1670
Db 2226992 CCGTTGGCGCGGCAACAGCGCTTCTCGAGTCCGAGCGCTCGCTGAAATGGGTGGCAAC 2226933
QY 1671 GCTTGTCTGAA-----ATCCGCTTAACCATCGCCATCAAGCAGATTCGATATCCACATGG 1726
Db 2226932 GCGTGTCTGAAACACACCGCGATCACCATTGCGGTAAACGCCGATTCATGGCGACATGG 2226873
QY 1727 TTTCTCTCCGCTTCAACGAGGTAGCTTTCTGGGCTGAGCAACCGCTCACGCTCGCTGTTG 1786
Db 2226872 TTTTCGCGCGTGTTCGAGC-----GTCTCGCGCAGCTCTCTCGAGTTCGGATC 2226822
QY 1787 GAAGATGAAGCGCACACATTTATCTTGTGCGGCTGGAGATGTTTATGAGCAGCGTAACC 1846
Db 2226821 GAGGACCAAGGACCATTCGCGCGCTGCTACCGGAGGCTGTGGCGATGGGCGCGTGAAC 2226762
QY 1847 CGTGAAGCTAATCCGCTGGCGGATGTGAGTAGTAGAATCTTGAACCATTCGCGCACCTGG 1906
Db 2226761 ACCGAGCGGAACCGCGTGCCTGGGCTGCGCGGTGACCCGCTGAAATCGCTACCTA 2226702
QY 1907 GCATTTGCAACCCCTCAATTGCGGATGCTTACATGTTGATGGGAACTAGATTGGGCT 1966
Db 2226701 CCAGTGGCCAGCAGCGCCATTCGTCAGGCGCATCTATCCGAGCGGTTCACTGCGCGCGG 2226642
QY 1967 GCGATGCCGCTTTACGCTTCGCTGCCAAAGATGTGCTTCAAGACCGGTGACCTGACCGGG 2026
Db 2226641 GCGGCTAAAGCTCCGCTCACTGGCGTGAATCGTGACGATGGGCTGCGAGGACATGTTGGTG 2226582
QY 2027 CGC---GTCCGATGCTGTTGGGCGCAGGCGGTATCCATTGTCTCCGTCGGCGGAAGT 2083
Db 2226581 CGTAAGCGCTTTTCGTCGCGCATCACCAGACCGCAGCATTTGTTCGCGACACAGAGGCGC 2226522
QY 2084 TTTGGTGAAGCAATTCGCGCGAGGCGCTTGGTGGGAGCTTCTTCCCGAAACCCAAAGCTGCT 2143
Db 2226521 TTCACCGCGCAGCGCGCGGCTGGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 2226462
QY 2144 CCCATGCTAAAGCAGGAGAGTGTATCTCTCTCGATGAGATACCCATTTGACACACCGATG 2203
Db 2226461 TCTCCGCTTGGCGATGATCGTTGCTAGCGGCTGTGCGACATACACCTCGACGCTCCCTCTC 2226402
QY 2204 TATTGGCAACGATGCGCGCTGGAACTAGATCTCTAGCTAGACTCACAGACCGCGCTGTT 2263
Db 2226401 TATTGGCAATGCTGGAACCTGGACATCCGATCATCGCGGAATTAACGACCGTGAAG 2226342
QY 2264 GATGAGCAATCGAGGATTCGCGCTTAGTTAC 2297
Db 2226341 GCGGCGCAAGCGCTGTGTACCGGCGCGCAAC 2226308
```


RESULT 3

US-09-103-840A-1/c
 ; Sequence 1, Application US/09103840A
 ; Patent No. 6294328
 ; GENERAL INFORMATION:
 ; APPLICANT: FLEISCHMAN, Robert D.
 ; APPLICANT: WHITE, Owen R.
 ; APPLICANT: FRASER, Claire M.
 ; APPLICANT: VENTER, John C.
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 ; FILE OF INVENTION: TUBERCULOSIS
 ; FILE REFERENCE: 24366-20007.00
 ; CURRENT APPLICATION NUMBER: US/09/103,840A
 ; CURRENT FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 4411529
 ; TYPE: DNA
 ; ORGANISM: Mycobacterium tuberculosis
 ; OTHER INFORMATION: H3RV
 US-09-103-840A-1

Query Match 5.1%; Score 121.2; DB 4; Length 4411529;
 Best Local Similarity 49.5%; Pred. No. 7.1e-24;
 Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;

QY 1433 CAACTGGACATTGCTCTCAATGATGAGGAGCGCTTGAAGCGCGCTCTTAGCC 1492
 DB 2229873 CAGCTGGCGCATGGCTGCCGTGCTGAACCTGCGAGCTTCGATGCGCGCGGAGCGC 2229814
 QY 1493 CTTTCAATTTCCCTCCGGGGTGAATCAGCGGTTAAAGCTCTCAGCATCAGTGGGT 1552
 DB 2229813 CTACATGTCACCCCTCGGCTGTGCTAGTCAGCGCATCAAGTCTGTGGAGCAGCAGGTGCGC 2229754
 QY 1553 CGAGTCTTGGTATCGGCGCAACCCGCGCAACGAGCGGTGAAGTCTCTTGTG 1612
 DB 2229753 CAGTGTCTGTGCTCAGGGAAGCATGTGCGGCGAGCGAGGTATCCCGTGTG 2229694
 QY 1613 CAAGCAGCGGGAATGTTGCTGCAAGCAGAACTAAAGCGCACTATCTGG--AC 1670
 DB 2229693 CGGTTGGCGCGCAACACGCTGCTGAGTCCGAGCGCTCGCTGAAATGGGTGGCAAC 2229634
 QY 1671 GCCTTCGTGAA---ATCCGTTAAACATCGCATCAACGAGATTCGTATCCACATGG 1726
 DB 2229633 CGCTCGTGAACGACGCGGATCACCATTGCGGTAAACGCGGATTCATGCGGACATGG 2229574
 QY 1727 TTTCTCTCCGTGTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTGCGCTTG 1786
 DB 2229573 TTTTCGCGCGTGTTCGACG-----GTCTCGGCGAGGTCTCTCGACGTTCCGATC 2229523
 QY 1787 GAAGATGAAGCGCACATATATCTTCTGCGCGGTGGAGATGTTTTAGAGCGGTAAAC 1846
 DB 2229522 GAGGACACGAGCAATTCGCGCGGCTGCTACGCGGAGGTGTGGCGATGGCGCGGTGACC 2229463
 QY 1847 CGTGAAGCTAATCCCTGCGGAGTGAAGTAGTAGAATCTTGAACCATGCGGCACCTTG 1906
 DB 2229462 ACCGAGCGGAACCGGTGCGGCGTCCCGGTGCAACCGCGTGGTGAATGCGCTACCTA 2229403
 QY 1907 GCATTTGAACCCCTCATTTGCGGATGCTATCATGTTGATGGGAACTAGATTGGCT 1966
 DB 2229402 CAGTGGCGCAGAGCGCATTCGTGAGCGGCAATCTATCGAGGGTTCACTGCGCGCGG 2229343
 QY 1967 GCGATCCCGCTTTACGCTTTCGTTCCCAAGATGTCCTTCAAGACCGGTGACCTGGACGGG 2026
 DB 2229342 GCGGTAAAGCTCCGTCACTGCGGTGAATCTGTAGACATGGCTGCGGACATGTTGTG 2229283
 QY 2027 CGC--GTGATGTTCTGTGGGGCGCAGCGGCTATTCATTTGTCCTGCGCGGAGGT 2083
 DB 2229282 CGTAGGCGCTTTCTGTCGCGCATCAACGAGCGCATTTGTCTCCGACCAACAGAGGGC 2229223

QY 2084 TTTGCTGAGGCAATTCGCGGAGGCGCTTGGTTGGGACTTCTTCCGAAACCCAGCTGCT 2143
 DB 2229222 TTCACCGCGCAGCGCGCGCGGCTGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 2229163
 QY 2144 CCATGCTAAAGCAGGAGAGTGAATCTCTCTCGATGAGATACCATTTGACACACCGATG 2203
 DB 2229162 TCTCCGCTTCCGATGATCGTTCTGTACGGGTCTGCGGACATACACCTCGACGTCCTCTC 2229103
 QY 2204 TATTGCGCAACGATGCGCGCTTGGAACTCTAGATCTCTAGCTAGACTACACAGCGCGTCTGTT 2263
 DB 2229102 TATTGGCAATGCTGGAACTGGACAGTCCGANTATCCGCGAATTTACCGACCGTGAGG 2229043
 QY 2264 GATGACGCAATTCGAGGATTCGCGCTTTAGTTAC 2297
 DB 2229042 GCGCGCGCAAGCGCTCTGTACCGGGCGCAGCAAC 2229009

RESULT 4

US-08-920-812-20/c
 ; Sequence 20, Application US/08920812
 ; Patent No. 5763188
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Teuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/920,812
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,577
 ; FILING DATE: 27-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5541 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli
 ; STRAIN: Clinical Isolate EC-625
 US-08-920-812-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;
 Best Local Similarity 50.5%; Pred. No. 1.7e-19;
 Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;
 QY 1 AGATACTCTTTTGAAGAAACCATGCTAGCATTTGGTTCACATTGTTGGTCTGGAAGGC 60

Db 3416 AATACCGCATGGAAGAACCGCTCTGCGTGGCTCATCGGTACAAAGCGGTAAAGC 3357
QY 61 TCTTTAGTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120
Db 3356 GCTTTATGTGCGGATCTCTTCTACTCGCCAGAGCGGACGCAAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCTGCGCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATCGTTG 180
Db 3296 GCGGAGTGAATAATTCGCTGTTAATTCATCAACCTTCGTACATTTACTGAACCGCTG 3237
QY 181 GGTGAGGAACCGGCGGATGACGGTGAGAACTTTGTCAGTCACTGCGCAACAAATGTTCT 240
Db 3236 GGTGGAATAAAGCGG-----CCTGCTGATACCTCTGACCGGAAATATCTCAACGG 3192
QY 241 TGGGCTATTGCTTTCTACCACTTGGGAGGCGCTGCTCACGCAAAATATCTCGATGG 300
Db 3191 GGGCTGATTGCGCTTACTCTCTGCTCAGGATTTGCTGACCGGAAATATCTCAACGG 3132
QY 301 AATTCCAGAGGGTTCCCG-----CGCCAGCGGGAAGTTCCTGTCAGGG 348
Db 3131 CATTCCGAGATTCAGGATTCACGGATGATCGTGAAGGAATAAAGTTCTGACCGCGAA 3072
QY 349 CATGTTGAACGTCGAACAAATATTGATGTTCCGCAAGCTCAATGACATCGCCAGGAACG 408
Db 3071 AATGCTTACCGAAGCCAACTCAACAGCTTACACTTATTGAATGAATGCGACAGCGG 3012
QY 409 CGGGCAGTCACTTGGCAGATGGCGTTGTCATGCGGTGCTCGCGGAGCAAGGAGTACGG 468
Db 3011 TGGACAATCAATGCGCAATGCGGTTAAGCTGTTGCTGAAAGA-----2967
QY 469 CGCGGATACCGTGAACAGTGCATTTGATGTTGCTTCAGTTCAGTTCGACAGTGAACAG 528
Db 2966 ----TGATCGGTGACGTCGGTATTGATTGTTGTC--CAGCGCGCGGACCACTTGAGGAGNA 2911
QY 529 CTTGATTCACTCAACAACTTGGAGTTTCTGACCGCGAGTTGGAGGCGATCGATGAGAT 588
Db 2910 CGTGAGGCGCTGAATAATCTGACATTTAGCACCGAGAGCTGGCGGAGTATGATCAGCA 2851
QY 589 TTCCACGACGCGCGCATCAACATTTTGGCGGAAGCCCGATTCCTCAAAACCCGCGAANA 648

RESULT 5

US-08-920-827-20/c
; Sequence 20, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneya
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5541 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-625
; US-08-920-827-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;
Best Local Similarity 50.5%; Pred. No. 1.7e-19;
Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;
QY 1 AGATACTCTTTGGAGAAACCATGTACGATTGCGTGACATTTGTTGCGTCTGGAAGCG 60
Db 3416 AATATACCGCATGGAAGAAACCGCTCTGCGTGGCTCATGCGGTACAAAGCGGTAAAGC 3357
QY 61 TCTTTAGTGGGTATTTCTTCTACGGTCCAGAGCTCACGCGGAGGCGGTGAGTTTCA 120
Db 3356 GCTTTATGTGCGGATCTCTCTTACTCGCCAGAGCGGACGCAAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCTCGCCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATTCGTTG 180
Db 3296 GCGGAGTGAATAATTCGCTGTTAATTCATCAACCTTCGTACATTTACTGAACCGCTG 3237
QY 181 GGTGAGGAACCGGCGGATGACGGTGAGAACTTTGTCAGTCACTGCGCAACAAATGTTCT 240
Db 3236 GGTGGAATAAAGCGG-----CCTGCTGATACCTCTGCAAAATAACGCGGT 3192
QY 241 TGGGCTCATTTGCTTTCTCACCACTTGGCAGGCGCTGCTCAGGCAAAATATCTCGATGG 300
Db 3191 GGGCTGATTGCTTTTACTCTCTCGCTCAGGATTTGCTGACCGGAAATATCTCAACGG 3132
QY 301 AATTCCAGAGGGTTCCCG-----CGCCAGCGGGAAGTTCCTGTCGTGAGGG 348
Db 3131 CATTCCGCAAGATTCACGGATTCATCGGATGATCGTGAAGGAATAAAGTTCTGACCGCGNA 3072
QY 349 CATGTTGAACGTCGAACAAATATTGATGTTCCGCAAGCTCAATGACATCGCCAGGAACG 408
Db 3071 AATGCTTACCGAAGCCAACTCAACAGCTTACACTTATTGAATGAATGCGACAGCGG 3012
QY 409 CGGGCAGTCACTTGGCAGATGGCGCTTTCATGCGGTGCTCGCGGAGCAAGGAGTACGG 468
Db 3011 TGGACAATCAATGCGCAATGCGGTTAAGCTGTTGCTGAAAGA-----2967
QY 469 CGCGGATACCGTGAACAGTGCATTTGATGTTGCTTCAGTTCAGTTCGACAGTGAACAG 528
Db 2966 ----TGATCGGTGACGTCGGTATTGATTGTTGTC--CAGCGCGCGGACCACTTGAGGAGNA 2911
QY 529 CTTGATTCACTCAACAACTTGGAGTTTCTGACCGCGAGTTGGAGGCGATCGATGAGAT 588
Db 2910 CGTGAGGCGCTGAATAATCTGACATTTAGCACCGAGAGCTGGCGGAGTATGATCAGCA 2851
QY 589 TTCCACGACGCGCGCATCAACATTTTGGCGGAAGCCCGATTCCTCAAAACCCGCGAANA 648

Db 2850 TATCGCGGATGGGAGTGAATCTGTGGCAGGCGTCTTCGGATAAATGACCTGTGTAATA 2791
 QY 649 CTAACCCATCAACATCAGTTGATGGCCATGCGGTTCATCAACTGCCACGACGAGTT 708
 Db 2790 CGGCGCGGAGAACCGCCCGCATAAGATCAGTGATATACGAGTCAGTCGGTCAAGAT 2731
 QY 709 GATCCAGCGCCACACCTTGGGGCTGGAGACGCGGCGGTGACAAATG 752
 Db 2730 AGCCCATCAAAATGACAGAGACAAACGTCAGATGGAATG 2687

RESULT 6

US-08-921-177-20/c
 ; Sequence 20, Application US/08921177
 ; Patent No. 5798211
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/921,177
 ; FILING DATE: 29-AUG-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/362,577
 ; FILING DATE: 27-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 19036/32420
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312/474-6300
 ; TELEFAX: 312/474-0448
 ; TELEX: 25-3856
 ; INFORMATION FOR SEQ ID NO: 20:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5541 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: Genomic DNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Escherichia coli
 ; STRAIN: Clinical Isolate EC-625
 ; US-08-921-177-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;
 Best Local Similarity 50.5%; Pred. No. 1.7e-19;
 Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;
 QY 1 AGATACTCTTTTGGAGAAACCATGTACGCAATGCGTGCATTTGGCTGTGGAAGGC 60
 Db 3416 AATACGCCGATGAGAAACCCCTCTGGCTGGCTCATGCGGTACAAAGCGGTAAAGC 3357
 QY 61 TCTTTAGTGGGATTTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120
 Db 3356 GCTTTATGCGGATCTCTCTTACTCGCCAGAGCGGACGCAAAAATGGTTGAGTTGCT 3297

QY 121 GSCGAGGAGGCTCCCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATCGTGT 180
 Db 3296 GCGCAGTGGAAAAATTCGCTGTTAAATTCATCAACCTTCGTACAATTTACTGAACCGCTG 3237
 QY 181 GGTGAGGAACCGGCGGATGACGGTGTGAGAACTTGTGTGAGTCAGCTGCAACAATGGTCT 240
 Db 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTGCAAAATACGGCGT 3192
 QY 241 TGGCGTCAATGCTTTCTCACCATTGCGCAGGCGCTGTCTACGGACAAATATCTCGATGG 300
 Db 3191 GGGCTGTATTGCTTTACTCTCTGCTCAGGGATTGCTGACCGGAAAATATCTCAACGG 3132
 QY 301 AATTCACAGAGGTTCCCG-----CGCCAGCCAGGGTAAGTCCCTGTCTGAGGG 348
 Db 3131 CATTCGCAAGATTCACGGATGATCGTGAAGGGAAATTAAGTTCTGTGCTGACCCCGAA 3072
 QY 349 CATGTTGAACGTGAACAATATTGATATATGTCGCGCAAGCTCAATGACATCGCCCGAAGACG 408
 Db 3071 AATGCTTACCGAAGCAACCTCAACAGCCTACACTTTATTGAATGAATGGCACAGCAGCG 3012
 QY 409 CGGCGAGTCACCTTGGCGAGATGGCGCTTGCATGGGTGCTGCGCGAGCAAGGAGAGTACGG 468
 Db 3011 TGGACAATCAATGGCGCAATGGCGCTTAAGCTGTGCTGAAAGA-----2967
 QY 469 CGCGGATACCGTGACAGTCGATTTGATTTGATGATGCTCGCAAGCTCAATGACATCGCCCGAAGACG 528
 Db 2966 ----TGATCGGTGACGTGCGTATTGATTTGTTGTC-CAGCCGCGGACCACTTGAGGAGAA 2911
 QY 529 CCTTGATTCACCAACAACCTTGGAGTTTCTGACGCCGAGTTGGAGGCGATCGATGAGAT 588
 Db 2910 CGTGAGGCGCTGATATCTGACATTTAGCACCGAGAGCTGGCGCAGATTTGATCACA 2851
 QY 589 TTCCACGACGCGCGCATCAACATTTGGCGGAAGCCACCGATTCCAAAACCCCGCAAAA 648
 Db 2850 TATCGCGATGGCGAGCTGAATCTGTGGCAGCGCTTCCGATAAATGACCTGTTTAATA 2791
 QY 649 CTAACCCATCAACATCAGTTTGTATGCGCCCAATGCGGTCTATCAACTGCCACGACGCTT 708
 Db 2790 CGGCGCGAGAACCGCGCCCGCATTAAGATCAGTGAATTAATGAGTGTGATGATGAT 2731
 QY 709 GATCCAGCGCCACACCTTGGGGCTGGACAGCGGCGGTGACAAATG 752
 Db 2730 AGCCCATCAAAATGACAGAGACAAACGTCAGATGGAATG 2687

RESULT 7

US-08-362-577C-20/c
 ; Sequence 20, Application US/08362577C
 ; Patent No. 5807673
 ; GENERAL INFORMATION:
 ; APPLICANT: Ohno, Tsuneya
 ; APPLICANT: Matsuhisa, Akio
 ; APPLICANT: Uehara, Hirotsugu
 ; APPLICANT: Eda, Soji
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606-6402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/362,577C
 ; FILING DATE: 27-MAR-1995
 ; CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 5541 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: Clinical Isolate EC-625
US-08-362-577C-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;
Best Local Similarity 50.5%; Pred. No. 1.7e-19;
Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;

QY 1 AGATACCTCTTTCGAGAAACCATGATGCGTATGCGTATGTTGGCTCTGGAAGGC 60
DB 3416 AATACGCGATGAGAAACCGCTCTGCGTGGCTATGCGGTACAAAGCGGTAGGC 3357
QY 61 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGCGGCTGAGTTTCA 120
DB 3356 GCTTTATGTGCGGATCTCTCTTACTCGCGAGCGGACCAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCGTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 180
DB 3296 GCGGAGTGGAAAAATTCGCTGTTAATTCATCACTTCGTACAATTTACTGACCGCTG 3237
QY 181 GGTGAGGAAACCGGGCGATGACGGTGAGAACTTTGTGCACTGAGTCGCAACAATGGTCT 240
DB 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTTGCAAAATAACGGCGT 3192
QY 241 TGGCGTATGTTTCTACCACTTGGCGAGGCGCTGCTCAGCGAACAATATCTCGATGG 300
DB 3191 GGGCTGTATGCTTTACTCTCTGCTCAGGATGCTGACCGGAAATATCTCAACGG 3132
QY 301 AATCCAGAGGTTCCG-----CGCCAGCGGTTAAGTCCCTGCTCAGGG 348
DB 3131 CATTCGCAAGATTTCAGGATGATCGTGAAGGGAATAAAGTTTCGTGCTCAGCGCGAA 3072
QY 349 CATGTTGAACGTGAACAATATGATATGTCGCAAGCTCAATGACATCGCCAGGAACG 408
DB 3071 AATGCTTACGAGGCAACCTCAACAGCCTACACTTATTTGAATGAATGGCACAGCG 3012
QY 409 CGGCGAGTCACTTTCGCGAGATGCGCTTGCATGGGTCTCGCGCAAGAGGAGTACGG 468
DB 3011 TGGACAATCAATGCGCAATGCGTTAAGCTGTTGCTGAAGA-----2967
QY 469 CGCGGATACGTCACCAAGTCATGTTGCTTGGTCTGTCAGTTGACAGCTGGAACAAG 528
DB 2966 ---TGATCGGTCAGCTCGGTATGATGTTGTTGTC-CAGCCGCGCGGCAACTTGAAGAGAA 2911
QY 529 CTTGATTTCACTCAACAACCTTGAGTGTCTGACCGCGAGTTTGGAGCGCATCGATCAGAT 588
DB 2910 CGTGCAGGCGCTGAATATCTGACATTTAGCACCGAGGAGCTGGCCGATGATCAGCA 2851
QY 589 TTCCGAGGCGCGCATCAACATTTTGGCGAAGGCGACCGATTCCAAACCGCGGAAA 648
DB 2850 TATCGCGGATGGCGAGCTGAATCTGTGGCAGGCGTCTTCGATAAATGACCTGTTAATA 2791
QY 649 CTAACCCATCAACATCAGTTTGTATGCCAATGCGGTTCATCACTGCCACGACGCTT 708
DB 2790 CGGGCGGAGAACCGGCCCGCATATAGATCAGTGTATGACAGATCGATCGGTCAAGAT 2731
QY 709 GATCCAGCGCCACACCTTGGGCTGGACAGCGGCGGTGACAATG 752

DB 2730 AGCCCATCAAAATGCGAGAGACAAACGTCAGATGATATG 2687

RESULT 8
US-08-920-828-20/c
Sequence 20, Application US/08920828
Patent No. 585398
GENERAL INFORMATION:
APPLICANT: Ohno, Teuneya
APPLICANT: Matsuhisa, Akio
APPLICANT: Uehara, Hirotsugu
APPLICANT: Eda, Soji
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/920,828
FILING DATE: 29-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,577
FILING DATE: 27-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 19036/32420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 5541 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: Clinical Isolate EC-625
US-08-920-828-20

Query Match 4.2%; Score 100.8; DB 2; Length 5541;
Best Local Similarity 50.5%; Pred. No. 1.7e-19;
Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;

QY 1 AGATACCTCTTTCGAGAAACCATGATGCGTATGCGTATGTTGGCTCTGGAAGGC 60
DB 3416 AATACGCGATGAGAAACCGCTCTGCGTGGCTATGCGGTACAAAGCGGTAGGC 3357
QY 61 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGGCGGTGAGTTTCA 120
DB 3356 GCTTTATGTGCGGATCTCTCTTACTCGCGAGCGGACCAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCGTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 180
DB 3296 GCGGAGTGGAAAAATTCGCTGTTAATTCATCACTTCGTACAATTTACTGACCGCTG 3237
QY 181 GGTGAGGAAACCGGGCGATGACGGTGAAGAACTTTGTGCACTGAGTCGCAACAATGGTCT 240
DB 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTTGCAAAATAACGGCGT 3192

Query Match 1.7%; Score 39.6; DB 2; Length 1935;
Best Local Similarity 49.1%; Pred. No. 0.14;
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
QY 520 GGACAAACAGCTTGATTCACTCAACAACTGGAGTTTCTGACGCGAGTTGGAGCGGAT 579
DB 1263 GTRACCGGATCGGTCCGCCATCAGTACGCTGTGTGGCCCTCAAGTGGAGTCTCT 1322
QY 580 CGATGAGATTTCCACGAGCGCCGCGATCAACATTTGGGGGAAGGCCACCGATTTCCAAAC 639
DB 1323 CGTGGTCATTGGCCATAGCTGTGCGTGCATCAGGGCGCTCTCTCCCTCCAGGAAC 1382
QY 640 CCGGAAACTAACCATCAATCAGTTGATGGCCATGCGGTATCACAACCTGCCAC 699
DB 1383 CCGCGAGCTTACACCTTCCTTCGAGGACTGGGTTAAGATCGGCTTCATTGCCAA 1442
QY 700 GACGAGCTTGATCAGCGCCGACACCTTGGGGCTG 733
DB 1443 GATGAGGTAAAGAGACGCGCTCGGTGCGC 1476

RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 1.6%; Score 37.6; DB 4; Length 4403765;
Best Local Similarity 47.5%; Pred. No. 34; Mismatches 124; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 695 GCACAGCAGCTTGATCCAGCGCCACACCTTGGGGCTGACAGCGGGGTGACAAATGCT 754
DB 720208 GCCATCGACGCGGTAAACCGTGCAACGGTGTGTGCGCGGACCGCGCGATCTCCCG 720267
QY 755 GCTGCGCGGAAACCCAGCGGAGACCATGAGCTTCCCGGAAACCGCGCGCGGCG 814
DB 720268 GCGCGGAAAGAAACCGGCGGAAATCCCGCCAGCAGGTCTCTCGATCGTCGACGCTG 720327
QY 815 AAATCCACCGTCCGGTGTGCGCGTATTGCGCGGACCGCGCGGATAAACACAAACGG 874
DB 720328 TGTGTCGTGACCCCGAACATTCGTGTCGCGCGCGGTGCAAGTGAACAGCGCCACCG 720387
QY 875 TCCAAATACGATTCGGGTTCAACACAGGTGACGACGATTCGCAATCAACATGGGCTT 930
DB 720388 ACCGGGGGCGCGGAGCTCCGCGCGCGCGCTCGACGCGCAGGCGGAGTCTT 720443

RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328

GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match 1.6%; Score 37.6; DB 4; Length 4411529;
Best Local Similarity 47.5%; Pred. No. 34; Mismatches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
QY 695 GCACAGCAGCTTGATCCAGCGCCACACCTTGGGGCTGACAGCGGGGTGACAAATGCT 754
DB 718775 GCCATCGACGCGGTAAACCGGTGCAACGGTGTGCGCGGACCGCGCGATCTCCCG 718834
QY 755 GCTGCGCGGAAACCCAGCGGAAACCATGAGGTTCGCGGAAACCGCGCGGCGCG 814
DB 718835 GCGGCGGAAAGAAACCGGCGGAAATCCCGCCAGCAGGTCTCTCGATCGTCGACGCTG 718894
QY 815 AAATCCACCGTCCGGTGTGCGCGTATTGCGCGGACCGCGCGGATAAACACAAACGG 874
DB 718895 TGTGTCGTGACCCCGAACATTCGTGTCGCGCGCGGTTCGAGGTGAACAGCAGCGCCACCG 718954
QY 875 TCCAAATACGATTCGGGTTCAACACAGGTTCAGCAGTTCGCAATCAACATGGGCTT 930
DB 718955 ACCGGGGGCGCGGAGCTCCGCGCGCGCGCTCGACGCGCAGGCGGAGTCTT 719010

RESULT 13
US-08-358-117-1/c
; Sequence 1, Application US/08358117
; Patent No. 5608147
; GENERAL INFORMATION:
; APPLICANT: Kaphammer, Bryan J.
; TITLE OF INVENTION: tfda Gene Selectable Markers in Plants and the
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steine, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,117
; FILING DATE: 15-DEC-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1405.0030001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 751..1611
US-08-358-117-1

Query Match 1.6%; Score 37; DB 1; Length 2058;

Best Local Similarity 53.9%; Pred. No. 0.83; Mismatches 0; Gaps 0;

Matches 76; Conservative 0; Indels 65; Length 2058;

QY 1433 CAACTGGACACTTGTCTCAATCATTTGATGAAGGAGCTTCAAGGCGGCTCTTTAGCC 1492

Db 530 CAGCTTCGCTATTTCTGCTCGCGGAGGAGGCAAGTCTCGGTCGCGCGCGCGG 471

QY 1493 CTTTCCATTTCCCTCGCGGCTGAGTCAAGCGGTTAAAGCTCTCAGCATCACTGGGT 1552

Db 470 CTGCATATTTCCAGGCCCGCTCAGCGACAGATTCAAGCGCTCGAACAGCATCTGGG 411

QY 1553 CGAGTGTGGTATCGGCACC 1573

Db 410 GTGTTGTTGTCGAGCGCAGC 390

RESULT 14

US-08-470-588-1/c

; Sequence 1, Application US/08470588

; Patent No. 6100446

; GENERAL INFORMATION:

; APPLICANT: STREBER, WOLFGANG R.

; APPLICANT: TIMMIS, KENNETH N.

; APPLICANT: ZENK, MEINHART H.

; TITLE OF INVENTION: MICROORGANISMS AND PLASMIDS FOR

; TITLE OF INVENTION: 2,4-DICHLOROPHENOXYACETIC ACID (2,4-D) MONOOXYGENASE

; TITLE OF INVENTION: FORMATION AND PROCESS FOR THE PRODUCTION OF THESE PLASMIDS

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP

; STREET: 745 FIFTH AVENUE

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: USA

; ZIP: 10151

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/470,588

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: LAWRENCE, WILLIAM F.

; REGISTRATION NUMBER: 28,029

; REFERENCE/DOCKET NUMBER: 514413-3526

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-588-0800

; TELEFAX: 212-588-0500

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2058 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-470-588-1

Query Match 1.6%; Score 37; DB 3; Length 2058;

Best Local Similarity 53.9%; Pred. No. 0.83;

Matches 76; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1433 CAACTGGACACTTGTCTCAATCATTTGATGAAGGAGCTTCAAGGCGGCTCTTTAGCC 1492

Db 530 CAGCTTCGCTATTTCTGCTCGCGGAGGAGGCAAGTCTCGGTCGCGCGCGCGG 471

QY 1493 CTTTCCATTTCCCTCGCGGCTGAGTCAAGCGGTTAAAGCTCTCAGCATCACTGGGT 1552

Db 470 CTGCATATTTCCAGGCCCGGTCAGCGACAGATTCAAGCGCTCGAACAGCATCTGGG 411

QY 1553 CGAGTGTGGTATCGGCACC 1573

Db 410 GTGTTGTTGTCGAGCGCAGC 390

RESULT 15

US-09-134-001C-928/c

; Sequence 928, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 928

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-928

Query Match 1.5%; Score 35.4; DB 4; Length 1149;

Best Local Similarity 51.6%; Pred. No. 1.8;

Matches 81; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1270 TTTAATCACCAGTACATTTCTGCGGTCGATGGACAGTAAAGACTGGCCCCCAAGCAG 1329

Db 934 TATTAGCTACAGCATATGTCAGCACACCATGTTTGTAGTGAAGTATCATGAGTAG 875

QY 1330 ACCTGTATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATG 1389

Db 874 TAACTCTATCAGTTGTTTCAAAAATTCACCTTGATCTATCGCAATATCTACAATCACTG 815

QY 1390 ATTGGTTCTTAACATGGTTTAAATATAGTTTCATGAAC 1426

Db 814 ATCCATCTTTTCATTTCTTTTATCATATCTTTCAGTCAC 778

Search completed: April 27, 2003, 12:21:20

Job time : 18178.8 secs

